The Dissertation Committee for Jaehyun Joo Certifies that this is the approved version of the following dissertation:

The Roles of Genetic Variation and Exercise on Dietary Patterns among Sedentary Young Adults

Committee:

Molly S. Bray, Supervisor

Jaimie N. Davis

Stefano Tiziani

Peter Mueller

Sinead A. Williamson

The Roles of Genetic Variation and Exercise on Dietary Patterns among Sedentary Young Adults

by

Jaehyun Joo

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2018

Dedicated to my family.

Acknowledgements

First and foremost, my deepest gratitude goes to my advisor, Molly Bray, for her guidance and constant support during my years as a Ph.D. student. Her enthusiasm and knowledge for the work she is doing has inspired my own throughout this journey.

My heartfelt appreciation goes to my committee members. I am deeply grateful to Sinead Williamson, who guided me with constructive ideas and practical advice for the technical details of this dissertation. I would like to thank Peter Mueller for invaluable conversations that inspired me to try cutting-edge statistical models. I also thank Jaimie Davis and Stefano Tiziani for their insightful comments and suggestions.

I have had the privilege of interacting and collaborating with a number of superb colleagues and would like to specifically mention Laura Lasinger, Fei Dong, Diana Gutierrez, Vasavi Shabrish, Meredith First, and Johanna Wright. Their insights, and discussions have enriched my ideas and heightened happy experience in my academic life.

I would like to give special thanks to Sungyeon Heo for being an inspiring colleague and a great friend.

Finally, I would like to thank my parents, Jongho Joo and Hyunwoo Lee, and brother Seokhoon Joo who have given me the endless support and care.

The Roles of Genetic Variation and Exercise on Dietary Patterns among Sedentary Young Adults

Jaehyun Joo, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Molly S. Bray

Researchers from health care to basic science are interested in determining how people can improve their overall eating patterns. This dissertation concerns the roles of genetic variation and exercise in influencing dietary patterns, with three specific aims.

1. Characterize dietary patterns using sparse latent factor models, addressing the methodological shortcomings of the current dietary pattern approach.
2. Examine the influence of a 15-week aerobic exercise training protocol on dietary patterns among young adults.
3. Examine the contribution of genetic variation to the variability of dietary patterns using genome-wide single nucleotide polymorphism (SNP) data.

Since foods and beverages are consumed in a variety of combinations under free-living circumstances, analysis performed on a per-food or per-nutrient basis can be misleading. Thus, it is important to consider not only research related individual dietary component, but also research that examines dietary patterns that represent the real-life

manner in which foods are consumed in a variety of combinations. With this idea in mind, dietary pattern analysis was performed to identify habitual dietary patterns among young adults. In this research project, sparse latent factor modeling was applied to dietary pattern analysis an alternative approach to overcome the methodological shortcomings of the current dietary pattern methods in which several steps are performed sequentially, and each step involves in an arbitrary decision rule.

Once dietary patterns were identified, the influence of exercise on patterns of food consumption was evaluated. Changes in dietary patterns among young adults after 15 weeks of exercise training were examined using the time-varying factor scores of dietary patterns. The results suggest that engaging in regular exercise may motivate individuals to regulate overall food intake as well as pursue healthier dietary preferences.

Finally, the SNP-based heritability of dietary patterns was examined using genome-wide SNP data. The proportion of variance in dietary patterns accounted for by genetic variation was calculated, which produced estimates of SNP-based heritability. The heritability of both the Snacking and Western patterns was greater than that of the Prudent pattern, suggesting a stronger biological underpinning for preference of foods high in fat and sugar.

Dietary patterns are more than the sum of their part; they represent what people habitually eat and drink in daily lives, and these dietary components act synergistically to affect health. Understanding the roles of exercise and genetic variation in influencing dietary patterns as a whole would be a significant step toward the goal of providing a

robust behavioral and biological basis for healthy eating patterns that both reduce risk of preventable chronic disease and promote overall health.

Table of Contents

List of Tables xi

List of Figures xii

Chapter 1. Introduction 1

Chapter 2. Literature Review 4

2.1. Dietary Pattern Analysis 4

2.2. Effects of Exercise on Dietary Intake 13

2.3. Genetics of Dietary Intake Phenotypes in Human 18

Chapter 3. Identifying Dietary Patterns among Young Adults using Sparse Latent Factor Models 32

Abstract 32

3.1 Introduction 34

3.2 Methods 37

3.3 Results 44

3.4 Discussion 54

Chapter 4. The Influence of 15-week Exercise Training on Dietary Preferences among Young Adults 63

Abstract 63

4.1 Introduction 65

4.2 Methods 68

4.3 Results 73

4.4 Discussion 83

Chapter 5. Genetics of Dietary Patterns among College Students 88

Abstract 88

5.1 Introduction 90

5.2 Methods 93

5.3 Results 97
5.4 Discussion 102

Chapter 6. Conclusion 108
Future Directions 114

Appendices 116
Appendix A. Sparsity-inducing priors for Bayesian analysis of sparse latent factor model used for deriving dietary patterns among the TIGER study participants 116
Appendix B. Posterior inclusion probability and factor loading estimate with its 95% credible interval 118
Appendix C. Model specifications for the estimation of factor scores 126

References 129

List of Tables

Table 4.1: Descriptive statistics of exercise characteristics and changes of anthropometric variables.....74

Table 5.1: Characteristics of the study subjects.....99

List of Figures

Figure 2.1: The scree test displays the eigenvalues in descending order of magnitude and determine where they appear to level off to the right of the plot to graphically select the number of meaningful factors 8

Figure 3.1: The standard interpretation of dietary patterns using PCA procedure interpret is performed by setting arbitrary thresholds, whereas sparse latent factor model uses the posterior inclusion probability that a food and a specific dietary pattern has non-zero association 44

Figure 3.2: The structure of latent factors is visualized using the heatmap where row correspond to food items and columns correspond to dietary patterns49

Figure 3.3: The simplified PCA-driven dietary patterns with the different values of cut-off on the magnitude of the rotated PCA loadings. 50

Figure 3.4: The qualitative comparison of food subsets for dietary patterns between PCA with Varimax rotation and the sparse latent factor analysis. ...51

Figure 3.5: The influences of gender, race/ethnicity, and BMI were simultaneously quantified via the sparse latent factor analysis 53

Figure 4.1: Dietary patterns among young adults identified in the TIGER study. (A) Posterior inclusion probabilities of individual food-factor pairs 75

Figure 4.2: Exercise treatment effect on dietary preferences based on changes in dietary pattern scores for all participants 77

Figure 4.3: Changes of individual food item intakes based on a multivariate linear mixed model.....80

Figure 4.4: Influences of exercise characteristics on the treatment effect of exercise intervention on dietary preferences.....81
Figure 4.5: Influence of gender and race/ethnicity on exercise treatment effect on dietary preferences.....82
Figure 5.1: Heritability estimates for habitual dietary patterns among college students with 95% Bayesian credible intervals.....100
Figure 5.2: Heritability estimates for food group intakes among college students with 95% Bayesian credible intervals.....101

Chapter 1. Introduction

A healthy diet plays a critical role in maintaining overall well-being. Today's obesogenic food environment, however, promotes unhealthy nutritional habits and increases the risk for preventable chronic diseases. It is, therefore, important to analyze the factors that drive specific eating behaviors to bring about effective dietary modifications that favor beneficial health outcomes.

Much work on this topic has focused on single nutrients and foods. Although this evidence base continues to be substantial, individually examining dietary components may fail to acknowledge complex interrelationships among foods and nutrients. In addition, simply linking a large number of isolated dietary measures to clinical covariates may make the inferred results less translatable to individuals and health professionals.

Latent factor models help to expose patterns in data by allowing the collection of observed variables to be modeled via a linear superposition of latent factors. In dietary pattern analysis, principal component analysis (PCA) and exploratory factor analysis (EFA) are the most popular of these models, and have been used to elucidate patterns of foods that individuals habitually eat and drink together. Nonetheless, numerous drawbacks of the current dietary pattern methods have been pointed out in the literature. A major limitation associated with both PCA and EFA is a series of arbitrary decisions involved at various steps of the

modeling effort. In this dissertation, a sparse variant of latent factor models was employed in a Bayesian context, to provide a more coherent procedure for identifying dietary patterns. The applicability and advantages of this approach were demonstrated with an empirical study in comparison with PCA.

With this dietary pattern modeling framework in hand, the influence of a 15-week exercise training protocol on dietary patterns among young adults was investigated. Most studies to date have centered on changes in single nutrients/foods in response to exercise. Nevertheless, modifying one dietary component is generally accompanied by changes in other aspects of diet, and therefore, the impact of exercise on dietary intake may be more complicated than previously described. Of key interest was to reveal the role of exercise on changes in overall dietary patterns, accommodating the complex interrelationships in an individual's total diet. Changes of dietary patterns in response to aerobic exercise training were analyzed in relation to different components of exercise characteristics, including duration, intensity, and dose.

Genetic association studies have identified a number of SNPs associated with food consumption. However, the functional contributions of individual SNPs have not been clearly elucidated and explain only a small proportion of total phenotypic variance. For this reason, an alternative approach was taken to infer the genetic underpinnings of a specific pattern of food intake. Specifically, the proportion of genetic variance for dietary patterns was estimated using a set of

SNPs collectively, which provides a means for accounting for far more of the genetic contribution to phenotypic variability than the small set of putatively casual SNPs found thus far.

This dissertation is organized as follows:

- Chapter 2 reviews the published literature related to the aims of this dissertation project.
- Chapter 3 describes the application of Bayesian sparse latent factor models to dietary pattern analysis and discusses the advantages and disadvantages of the proposed framework, in comparison with the current dietary pattern approach.
- Chapter 4 presents the analysis of changes in dietary patterns in response to a 15-week exercise training protocol and addresses how these changes are associated with different components of exercise characteristics, including intensity, duration, and dose.
- Chapter 5 describes the SNP-based heritability of dietary patterns, adhering to the assumption that dietary phenotypes are dependent on the cumulative action of many small-effect genes. Keeping this in mind, the SNP-based heritability estimates for dietary patterns using all common SNPs simultaneously were examined and interpreted.
- Finally, Chapter 6 provides a summary of this dissertation and directions for future work.

Chapter 2. Literature Review

2.1. Dietary Pattern Analysis

In recent years, dietary pattern analysis has been used to represent what individuals habitually eat and drink (Hu, 2002). Dietary pattern analysis aims to address complex interrelationships among nutrients and foods, assuming a relatively small number of underlying patterns of dietary intake. Dietary patterns can be assessed in several ways, such as clustering, latent factor analysis, and reduced rank regression (Hoffmann, Schulze, Schienkiewitz, Nöthlings, & Boeing, 2004; Moeller et al., 2007; Newby & Tucker, 2004; Reedy et al., 2010). This dissertation focuses on latent factor analysis, which is the most frequently used method to find dietary patterns.

2.1.1 Latent Factor Models

Latent factor models aim to explain observed data in terms of a linear superposition of latent factors. In the context of a dietary pattern, these factors represent a set of food combinations that are habitually consumed together, which is are not directly observable. The traditional setup for factor analysis consists of $x_i = (x_{1,i}, \ldots, x_{p,i})'$, a p-vector of the observed intake of food $f = 1{:}\,p$ on sample $i = 1{:}\,n$. For a fixed dimension k, the model is of the form

$$x_i = A\lambda_i + \nu_i$$

where A is the factor loading matrix, $\lambda_i = (\lambda_{1,i}, \ldots, \lambda_{k,i})'$ is the k-vector of latent factor scores, and $\nu_i = (\nu_{1,i}, \ldots, \nu_{p,i})'$ is a p-vector of noise. The associations between each food and each of the underlying dietary patterns are indicated via the inference on the factor loading matrix A. The activity of each latent factor is expressed by the latent factor score λ_i. In the field of nutritional epidemiology, the most popular of these models has been principal component analysis (PCA) (Pearson, 1901) and exploratory factor analysis (EFA) (Spearman, 1904). As a non-probabilistic model, PCA does not explicitly model ν_i, which may lead to the confounding of noise due to its inherent empirical estimation of latent factors. In contrast, EFA aims to decouple these noise components from the assessment of the latent factor structure by assuming that ν_i is a vector of idiosyncratic, variable-specific noise (Spearman, 1904).

2.1.2 Procedure of Dietary Pattern Analysis

Currently, the standard PCA/EFA-based procedure for assessing dietary patterns consists of several sequential steps (Martinez, Marshall, & Sechrest, 1998; Moeller et al., 2007). For entry the analysis, measured intakes of food items are typically aggregated to a smaller number of food groups (Moeller et al., 2007; Newby & Tucker, 2004). These aggregated food groups are further processed to better approximate normal distributions because the distributions of food intake

for many food items are typically positively skewed (Theobald, Chatterjee, & Horgan, 2012).

After the construction of input variables, factor analysis of dietary patterns proceeds in four separate steps: (i) choosing the number of the latent factors (i.e., dietary patterns); (ii) estimation of factor loadings from a latent factor model; (iii) factor rotation to search for a simple structure; (iv) loading truncation to facilitate interpretation.

Once the structure of latent factors is determined, each factor is annotated according to its associations with food variables. The annotated factors (i.e., dietary patterns) are then used to summarize, organize, and understand the dietary data. Factor scores are also calculated based on the estimates of factor loadings. These scores are generally used as explanatory variables for studying associations between dietary patterns and health outcomes of interest in future analyses.

2.1.3 Methodological Issues in the Current Dietary Pattern Analysis

The major limitation of conventional dietary pattern analysis methods, PCA and EFA, is that the various steps are involved in a certain degree of arbitrariness, which may have non-negligible influence on the estimated factor structure (Browne, 2001; Martinez et al., 1998; Newby & Tucker, 2004).

One of key decisions in dietary pattern analysis is how to collapse measured intake of food items into food groups (Martinez et al., 1998; Newby &

Tucker, 2004). Collapsing food items may be useful in dealing with multi-modality of food intake distributions, which are generally a mixture of positively skewed distributions for consumers and point mass at zero for non-consumers. However, this process depends on judgmental decisions by researchers since there are numerous ways of aggregating food items into different numbers of food groups (Martinez et al., 1998). Indeed, food-grouping strategies have not been consistent across studies, although researchers may have considered how other studies have previously defined food groups (Newby & Tucker, 2004).

Both PCA and EFA assume that the number of latent factor are known in advance. Various criteria have been proposed to determine the number of latent factors to retain in the model. In dietary pattern analysis, Kaiser's rule (Kaiser, 1960) and Cattell's scree test (Raymond B. Cattell, 1966) have been the most utilized methods in practice (Moeller et al., 2007). Kaiser's rule computes the eigenvalues for the correlation matrix of input variables and keeps the factors that have eigenvalues greater than one. Cattell's scree test (Figure 2.1), on the other hand, plots the eigenvalues of the correlation matrix in descending order and finds the place where these values level off to the right of the graph to select the number of latent factors. However, despite their widespread use, both Kaiser's rule and Cattell's scree test, have long been criticized due to their ambiguity and subjectivity (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

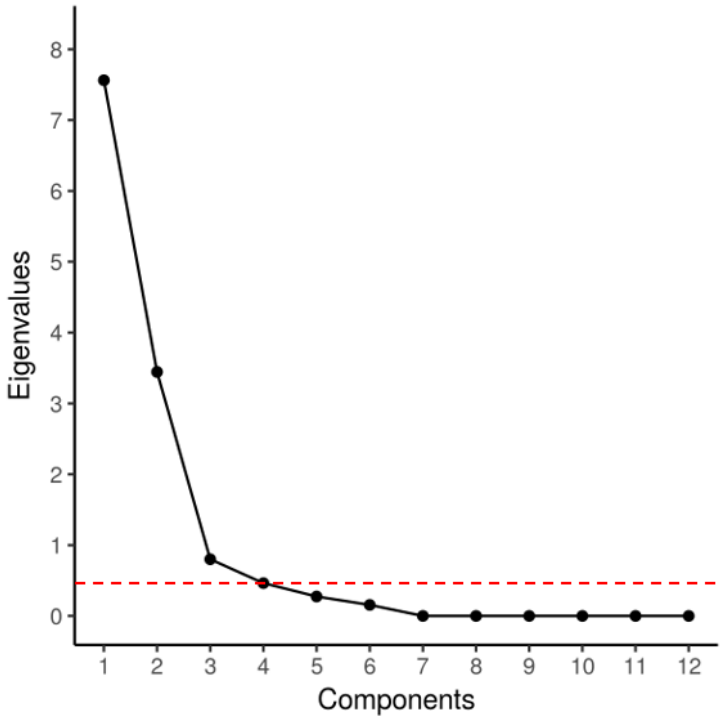


Figure 2.1: The scree test displays the eigenvalues in descending order of magnitude and determine where they appear to level off to the right of the plot to graphically select the number of meaningful factors. However, picking the "elbow" point can be subjective.

The traditional approach for obtaining interpretable patterns of factor loadings have entailed *post-hoc* rotations of the original estimates and loading truncations. It has been well known that traditional factor analysis can estimate latent factors and factor loadings only up to an orthogonal transformation, since

$$x_i = A\lambda_i + \nu_i = APP'\lambda_i + \nu_i = (AP)(P'\lambda_i) + \nu_i = A^*\lambda_i^* + \nu_i$$

for any orthogonal matrix P that satisfies $PP' = I$ if we set $A^* = AP$ and $\lambda_i^* = P'\lambda_i$. Hence, there is an infinite number of solutions that explain the implied

covariance structure in the data equally. Factor rotation is motivated to find a single solution that can simplify the factor structure and, thus, facilitate its interpretation (Raymond Bernard Cattell, 1978; Thurstone, 1947). Loosely speaking, a simple structure is achieved when the elements of the factor loading matrix have high contrast and many zeros so that each input variable is strongly associated with only one factor (Kano, Miyamoto, & Shimizu, 2003). Several factor rotation methods have been developed to accomplish this structure. Although the varimax rotation (Kaiser, 1958) is most common in dietary pattern analysis, the choice of rotation method is somewhat arbitrary due to the lack of a clear definition of a "simple" structure (Martinez et al., 1998). In addition, how the choice of a rotation method influences interpretation of dietary patterns has not been addressed well in the literature.

The obvious drawback of standard PCA and EFA approaches is that the factor loadings are typically non-zero, and therefore all input variables appear in all latent factors (Rasmussen & Bro, 2012; Zou, Hastie, & Tibshirani, 2006); Thus, it is often difficult to evaluate which food variables are most relevant to a given factor in interpreting the results. In dietary pattern analysis, it is a common practice to reduce the number of explicitly used food variables in describing latent factors by artificially setting the factor loadings with absolute values smaller than an arbitrary threshold to zero (Zou et al., 2006). In most studies, a threshold is typically chosen based on interpretation of the latent factors, but there is no

objective criteria that would explicitly define interpretability (Cadima & Jolliffe, 2001).

Other shortcomings of the current approaches include the lack of a rigorous strategy for dealing with missing values and limited flexibility to incorporate covariate information, such as gender, age, and cultural background, known or hypothesized to influence the dietary patterns.

2.1.4 Bayesian Sparsity and Latent Factor Models

The concept of sparsity has received attention as a mean of achieving simpler, and interpretable solutions in the field of statistics and machine learning (Rasmussen & Bro, 2012). A sparse statistical model assumes that only a small proportion of parameters will be non-zero. Sparsity can be one way of dealing with the rotational invariance in ordinary latent factor models via regularization that prioritizes a factor modeling matrix with zeros using a penalty or sparsity prior (Carvalho et al., 2008; Gao, Brown, & Engelhardt, 2013; Ročková & George, 2016). In the Bayesian context, sparse latent factor models have taken advantage of a sparsity-inducing prior. This prior in general has substantial probability mass around zero to provide strong shrinkage to zero, but also has heavy tails on plausible values to allow important factor loadings to escape shrinkage (Carvalho, Polson, & Scott, 2009; Lucas et al., 2006; Scott & Berger, 2006; West, 2003). Lucas et al. (Lucas et al., 2006) proposed the use of a

hierarchical sparsity prior on each element of the factor loading matrix, which induce sparsity using a two-component mixture model inducing a point mass at zero and a normal distribution. The prior takes the form

$$\alpha_{f,j} \sim (1-\pi_{f,j})\delta_0(\alpha_{f,j}) + \pi_{i,j}N(\alpha_{f,j} \mid 0, \tau_j)$$
$$\pi_{f,j} \sim (1-\rho_j)\delta_0(\pi_{f,j}) + \rho_j Be(\pi_{f,j} \mid a_j m_j, a_j(1-m_j))$$

where $\alpha_{f,j}$ is an element of the factor loading matrix; $\pi_{f,j}$ is the probability of $\alpha_{f,j}$ being non-zero; $\delta_0(\cdot)$ is a point-mass at zero; $Be(\pi_{f,j} \mid a_j m_j, a_j(1-m_j))$ is a beta distribution with factor-specific mean m_j and precision parameter $a_j > 0$; By integrating out the variable-specific probabilities $\pi_{f,j}$ from the prior for $\alpha_{f,j}$, $\pi_{f,j}$ can be simply replaced by $E(\pi_{f,j} \mid \rho_j) = \rho_j m_j$, which represents the common base-rate of non-zero factor loadings. This hierarchical prior naturally shrinks the posterior probability $\hat{\pi}_{f,j} = \Pr(\alpha_{f,j} \neq 0 \mid x_{1:n})$ towards zero for variables showing little evidence of association with j-th factor, while shrinking the non-zero $\hat{\pi}_{f,j}$ towards the estimated base-rate. The posterior probability $\hat{\pi}_{f,j}$ allows a probabilistic assessment of variable f being associated with j-th factor. Several alternatives may also be considered. A soft approach to incorporating sparsity is to place a Gamma prior on the precision of each element of the factor loading matrix, resulting in the elements being marginally student-t distributed (J. Geweke, 1993; Tipping, 2001). Such a construction does not impose exact zeros

on factor loadings, but many of them will be negligibly small. Another interesting construction is the use of the Indian Buffet Process (IBP), which defines a distribution over infinite binary matrices, in order to provide sparsity and define a generative process for the number of latent factors (Ghahramani & Griffiths, 2005; Knowles & Ghahramani, 2011). The basic form is

$$\Pr\left(\alpha_{f,j} \mid Z_{f,j}, \tau_j\right) \sim \left(1 - Z_{f,j}\right)\delta_0\left(\alpha_{f,j}\right) + Z_{f,j}N\left(\alpha_{f,j};\ 0, \tau_j^{-1}\right)$$

where Z is a binary matrix whose (f, j)-th element $Z_{f,j}$ represents whether observed variable f is associated with j-th factor; τ_j is the precision of the j-th factor; $\delta_0(\cdot)$ is a point-mass at zero; The IBP functions as an infinite-capacity prior that allows an unbounded number of latent factors, although only a finite number will have non-zero loadings (Knowles & Ghahramani, 2011).

Sparsity can be a useful add-on to dietary pattern analysis by providing an alternative approach to derive dietary patterns that include only a subset of food variables, instead of using a *post hoc* factor rotation and a threshold-based loading truncation. In chapter 3, the hierarchical sparsity-inducing priors proposed by Lucas et al. (Lucas et al., 2006) will be applied to dietary pattern analysis, in order to overcome several methodological shortcomings of the current dietary pattern methods.

2.2. Effects of Exercise on Dietary Intake

Exercise has the potential to regulate food intake in order to compensate the energy expended through physical activity (Blundell, Gibbons, Caudwell, Finlayson, & Hopkins, 2015; Horsch et al., 2015; Panek, Jones, & Temple, 2014). Most studies have focused on the effect of exercise on total energy intake and energy balance (Schubert, Desbrow, Sabapathy, & Leveritt, 2013), although exercise-induced health benefits are not necessarily dependent on changes in body weight or adiposity (King, Hopkins, Caudwell, Stubbs, & Blundell, 2009). Several cross-sectional studies have presented that physical activity was positively associated with consumption of healthy foods (Ax et al., 2016; Eaton et al., 1995; Oppert et al., 2006). Household survey data collected from southeastern New England communities showed that physically active individuals consumed more fiber, fruits, and vegetables, and ate less total fat (Eaton et al., 1995). In French adults, leisure-time and occupational physical activity were positively associated with increased frequency of fruits and vegetable consumption (Oppert et al., 2006). In Swedish adults, physical activity was positively associated with vegetables, fruits and seafood and inversely associated with foods high in sugar and fat (Oppert et al., 2006). In addition, Emmons et al. (Emmons et al., 2005) reported that physical activity promotes individuals' self-efficacy and motivation to modify diet. Similarly, Wilcox et al. (Wilcox, King, Castro, & Bortz, 2000) reported that participation in clinical trials of physical activity could result in

adoption of healthy diets beyond just the behavior targeted by intervention. Based on these data, one can formulate a hypothesis that exercise induces healthy dietary modification to counteract obesity, but a prospective relationship remains understudied.

2.2.1 Energy Intake

A fundamental principle of nutrition and metabolism is that body weight change results from an imbalance between energy intake and energy expenditure (Hall et al., 2012). It was previously thought that individuals show a compensatory increase in energy intake in response to exercise training; however, most studies of short exercise duration (< 2 weeks) showed little effect on energy intake (Blundell et al., 2015; King, Lluch, Stubbs, & Blundell, 1997; C. Martins, Morgan, & Truby, 2008). Conversely, several studies have reported a transient decrease in hunger after intense exercise, which is known as exercise-induced anorexia (Blundell, Stubbs, Hughes, Whybrow, & King, 2003; Broom, Stensel, Bishop, Burns, & Miyashita, 2007; Pomerleau, Imbeault, Parker, & Doucet, 2004), but such change was not consistently associated with a decrease in overall food intake (Schubert et al., 2013).

Longer studies (> 2 weeks) have demonstrated a partial compensation of energy intake in both men and women, which accounted for approximately 30% of the energy cost associated with exercise (Stubbs et al., 2002; Whybrow et al.,

2008). Thus, exercise can produce a negative energy balance and some degree of weight loss, though exercise alone has a small effect on body weight (Laskowski, 2012; K. Shaw, Gennat, O'Rourke, & Del Mar, 2006). Additionally, a recent systematic review found that there were significant limitations in the ability to make definite conclusions regarding the effects of exercise on dietary intake primarily due to methodological differences and statistically low power across studies (Donnelly et al., 2014).

2.2.2 Macronutrient Intake

Irrespective of study design, there has been limited reliable evidence that exercise changes macronutrient intake and composition (Donnelly et al., 2014; Washburn et al., 2015). To date, only a few studies observed a significant influence of exercise on absolute and/or composition of macronutrient intake (Donnelly et al., 2014).

Farah et al. (Farah, Malkova, & Gill, 2010) reported that carbohydrate and protein intake at lunch were higher in a 3-day exercise group than the no-exercise control group, although there was no difference in macronutrient intake at breakfast. Stubbs et al. (Stubbs et al., 2002) reported an elevation of fat and carbohydrate intake following a 7-day exercise regimen. A 16-week brisk walking intervention was associated with an increased in carbohydrate intake in sedentary

and obese white women, although this relationship was not observed in African Americans (Brandon & Elliott-Lloyd, 2006).

Conversely, some studies have shown decreases in the absolute intake of macronutrients (Bales et al., 2012; B. S. Shaw, Shaw, & Mamen, 2010). In males, 16-weeks of resistant training was associated with decreases in daily protein, carbohydrate, and fat intake, while endurance training was only associated with a decrease in fat intake (B. S. Shaw et al., 2010). In the STRRIDE AT/RT trial, individuals who received an 8-month aerobic exercise training intervention showed decreases in fat and protein intake, and those who performed resistance exercise training showed a decrease in carbohydrate intake (Bales et al., 2012).

In a 12-week controlled intervention study, female participants who were assigned to the 60 min/day brisk walking group increased the percentage energy from fat by 4.0%, with no significant percent change in carbohydrate and protein (Kirkwood, Aldujaili, & Drummond, 2007). In the Oslo Diet and Exercise Study, the exercise group showed a lower percentage of fat intake, compared to the control group (Reseland et al., 2001).

2.2.3 Food Preference

In humans, a limited number of studies have examined exercise-induced changes in food preference with isolated categories of foods and nutrients, in particular, preferences for foods with varying levels of salt and fat (Kanarek, Ryu,

& Przypek, 1995). An acute session of exercise was associated with an increase in the preference for salt, which might be due to sodium loss (Leshem, Abutbul, & Eilon, 1999). In addition, exercise was associated with a decrease in the relative preference for high fat vs. low fat foods, regardless of exercise modality (McNeil, Cadieux, Finlayson, Blundell, & Doucet, 2015). Similarly, short-term aerobic exercise (< 2 weeks) altered a motivation to consume high energy density foods in sedentary adults (Panek et al., 2014). Brisk walking was also shown to reduce the urge to consume sugary snacks and chocolate (Taylor & Oliver, 2009; Thayer, Peters, Takahashi, & Birkhead-Flight, 1993).

In animal studies, wheel running resulted in a reduction in the preference for a high-fat diet (Liang, Bello, & Moran, 2015; Scarpace, Matheny, & Zhang, 2010; T. Yang, Xu, York, & Liang, 2017). When rodents were presented with access to both normal chow and high-fat diet, they preferred the high-fat diet to normal chow in the sedentary condition. With wheel running access, rodents reduced the preference for the high-fat diet significantly (Liang et al., 2015; Scarpace et al., 2010; T. Yang et al., 2017). In obese mice, eight weeks of treadmill exercise led to decreases in the preference for a high-fat diet, along with increases in the preferences for sucrose and milk (Chen et al., 2017). With four weeks of swimming exercise, male mice showed a reduced preference for a high-fat diet and a slower rate of weight gain (H.-J. Wang, Yang, & Chen, 2017).

Although both human and animal studies support that exercise has a potential to alter not only energy intake but also food preferences, most studies to date have examined the influence of exercise on isolated categories of diet or single nutrients. Thus, of key interest in chapter 4 is to examine changes of overall dietary intake in response to exercise using dietary patterns, accounting for complex interrelationships of multiple dietary components.

2.3. Genetics of Dietary Intake Phenotypes in Human

Genetic differences may predispose certain individuals more than others to unhealthy dietary habits, such as increased preferences for energy-dense, and high fat foods (Rankinen & Bouchard, 2006). Family and twin studies have provided compelling evidence that multiple dietary intake phenotypes, such as energy and macronutrient intake, are clearly heritable (Grimm & Steinle, 2011; Rankinen & Bouchard, 2006). In addition, recent progress in genomics technologies has led to the identification of genes related to specific dietary behaviors via genetic association studies (Grimm & Steinle, 2011). Several studies have identified genes associated with patterns of dietary intake (Krom et al., 2007; J. M. McCaffery et al., 2017; Melhorn et al., 2018). However, the small contributions of each identified SNP to overall heritability limit our ability to robustly identify individuals who are at risk for unhealthy food intake phenotypes.

2.3.1 Genetics of Obesity and the Satiety Signaling

Human food intake is not an entirely voluntary process, but instead, is also driven by underlying biological processes (Farooqi & O'Rahilly, 2006). It has been well established that almost all forms of human monogenic obesity, which displays a Mendelian pattern of inheritance, have hyperphagia, or uncontrolled eating, as a common feature (Hu, 2008; O'rahilly & Farooqi, 2008). Genetic mutations responsible for monogenic obesity identified to date have been shown participate in a complex and highly redundant set of pathways that play a pivotal role in the neural control of food intake.

Leptin is secreted by adipose tissue in proportion to the amount of stored fat in the body (Zhang et al., 1994). In the brain, leptin protein binds to the leptin located on the surface of neurons of the arcuate nucleus. The downstream signals are involved in both catabolic and anabolic pathways, each consisting of distinct classes of neurons (Hu, 2008; Huvenne & Dubern, 2014; Oswal & Yeo, 2007). The catabolic pathway includes pro-opiomelanocortin (POMC) and the cocaine- and amphetamine-regulated transcript (CART), which reduce appetite and food intake via activation of the melanocortin 4 receptor (MC4R). The anabolic pathway includes the neuropeptide-Y (NPY) and agouti-related protein (AGRP), which increase appetite and food intake by reducing MC4R signaling.

The mutations in genes coding for leptin (*LEP*) and leptin receptor (*LEPR*) have been associated with rapid and dramatic increase in weight from the first

month of life (Clément et al., 1998; Fatima et al., 2011; Montague et al., 1997, 1997; Strosberg & Issad, 1999). In children from a consanguineous family of Pakistani origin, a *LEP* gene mutation was associated with early-onset severe obesity and hyperphagia (Montague et al., 1997). *LEPR* mutation carriers showed severe obesity, with a large amount of total body fat mass (> 50%) (Hu, 2008; Huvenne & Dubern, 2014). In addition, Krude et al. (Krude et al., 1998) reported that mutations in the *POMC* gene led to a defect in POMC function, resulting in severe obesity, alteration in pigmentation, and adrenocorticotropic hormone (ACTH) deficiency. A frameshift mutation in *MC4R* gene was associated with dominantly inherited obesity(Vaisse, Clement, Guy-Grand, & Froguel, 1998; Yeo et al., 1998). Later studies revealed that multiple common (allele frequency ~5%) mutations in *MC4R* are also associated with severe forms of obesity and food intake (Loos et al., 2008; Stutzmann et al., 2009; Valladares et al., 2010). Additional genes associated with monogenic forms of obesity and hyperphagia include *neurotropic tyrosine kinase receptor type 2* (*NTRK2*), *brain-derived neurotrophic factor* (*BDNF*), and *single-minded homologue 1* (*SIM 1*) (Saeed et al., 2018).

Studies of monogenic forms of obesity provide evidence that food intake is a heritable phenotype, and genetic variations in the leptin/melanocortin axis may have a potential to influence a variety of aspects of food intake. While complete disruption of satiety signaling pathway genes results in profound

hyperphagia and extreme forms of obesity, additional studies (described below) support the hypothesis that multiple genes, each with small effects, likely influence common patterns and preference for food intake.

Several family-based studies have examined the genetic basis of eating behavior using the Three-Factor Eating Questionnaire (TFEQ) (Stunkard & Messick, 1985). TFEQ measures three distinct traits: cognitive restraint, disinhibition, and susceptibility to hunger. In the Amish Family Study, heritability was estimated to be 0.28 for cognitive restraint, 0.40 for disinhibition, and 0.23 for hunger (Steinle et al., 2002). In the Quebec Family study, the estimated heritability was 0.06 for cognitive restraint, 0.18 for disinhibition, and 0.28 for hunger (Provencher et al., 2005). In the Virginia Twin Registry, the estimated heritability was 0.45 for disinhibition and 0.8 for hunger, but cognitive restraint had an insignificant value of heritability (Neale, Mazzeo, & Bulik, 2003).

Evidence supports that genetic variation in fat mass and obesity-associated gene (*FTO*) contributes obesity risk by influence satiety and hunger (Grimm & Steinle, 2011). Hoed et al.(den Hoed, Westerterp-Plantenga, Bouwman, Mariman, & Westerterp, 2009) reported that the A allele of rs9939609 SNP in *FTO* was associated with reduced postprandial responses in hunger and satiety, and its effect appeared to be mediated by an epistatic interaction involving variants in *LEPR* and DNA *methyltransferase 3 beta* gene (*DNMT3B*). In Scottish children, the TA/AA genotype of rs9939609 SNP in *FTO* was linked with increased energy

intake as a result of the increased intake of energy-dense foods (Cecil, Tavendale, Watt, Hetherington, & Palmer, 2008). In another study, children with TA/AA genotype were associated with more frequent "loss of control" eating episodes (e.g., binge eating) (Tanofsky-Kraff et al., 2009).

In addition, studies on genetic variants in digestive neuroendocrine hormones, such as ghrelin and leptin, provided insight into how their hormones and their genetic variants are associated with eating behavior (Grimm & Steinle, 2011). Ghrelin is a peptide hormone that promotes hunger and food intake through its receptor in the hypothalamus (Nakazato et al., 2001). In female Caucasian individuals, the Leu72Met polymorphism of the ghrelin gene, known as a risk variant for obesity and metabolic syndrome (Hinney et al., 2002; Korbonits et al., 2002; Steinle, Pollin, O'Connell, Mitchell, & Shuldiner, 2005), was associated with the prevalence of binge eating disorder (Monteleone, Tortorella, Castaldo, Di Filippo, & Maj, 2007).

Leptin and cholecystokinin work in opposition to ghrelin to stimulate satiety and have been implicated in the control of meal size and meal frequency (Berthoud, Sutton, Townsend, Patterson, & Zheng, 2006; Zorrilla, Inoue, Valdez, Tabarin, & Koob, 2005). In the EPIC study, genetic variants at *cholecystokinin* gene (*CCK*) and haplotypes were associated with extreme meal size (Krom et al., 2007). In this study, common variant at *LEP* was associated with extreme snacking behavior, but not with increased meal size.

2.3.2 Heritability of Dietary Intake

Heritability is defined as the proportion of the phenotypic variance that is attributable to genetic variation in a given population, taken at a particular time or age (J. Yang, Zeng, Goddard, Wray, & Visscher, 2017). Heritability is a key parameter in evaluating the efficiency of genetic risk prediction for a given trait and comparing the contribution of genetic factors on the same trait across populations and of different traits within a population (Visscher, Hill, & Wray, 2008). An important caveat is that heritability does not tell us about the magnitude of genetic variance, since it is a ratio of variances.

Heritability of various dietary traits has been reported from family-based studies (Pallister, Spector, & Menni, 2014; Rankinen & Bouchard, 2006). Total energy and macronutrient intakes, thus far, have been the most studied dietary traits, with studies reporting a sizable contribution of genetic variation to the variability for those phenotypes (Chmurzynska & Mlodzik, 2017; Pallister et al., 2014; Rankinen & Bouchard, 2006). The estimated heritability of energy intake ranged from 0.65 in 390 individuals (mean age 38.8 y) from the Minnesota Twin Registry (de Castro, 1993) to approximately half that value in 1,212 individuals (mean age 38 y) from the Danish Twin Registry in which the heritability in males was estimated to be 0.38 and in females 0.32 (Hasselbalch, Heitmann, Kyvik, & Sørensen, 2008). The heritability estimates of macronutrient intakes varied from 0.25 to 0.67 for carbohydrate, from 0.28 to 0.70 for protein, and from 0.35 to 0.54

for fat (de Castro, 1993; Hasselbalch et al., 2008; Hur, Bouchard, & Eckert, 1998; Pallister et al., 2014; Wade, Milner, & Krondl, 1981).

The heritability of energy and macronutrient intake appear to be dependent on age (Pallister et al., 2014). The relative importance of genetic variation increased with age, with genetic variance accounting for only about one tenth of the total variation for dietary phenotypes in twins aged 21 months from the Gemini twin birth cohort (Pimpin et al., 2013), whereas it explained a higher proportion, ranging from 0.31 for fat intake to 0.48 for total energy intake, in twins aged 11-13 years from the University of Southern California Twin study (Liu, Tuvblad, Raine, & Baker, 2013). The lower heritability of pre-school children could be explained by their reliance on immediate family for food in which parental intake and feeding styles, the availability of food in the home, or the foods offered to the children are the principal drivers of food consumption (McGowan, Croker, Wardle, & Cooke, 2012; Pimpin et al., 2013) .

Recently, several studies have investigated the heritability for individual foods or food groups (Pallister et al., 2014); however, the heritability of individual food consumption should be evaluated with a great deal of caution since there will be no capacity for food intake to be genetically influenced if food availability is limited in a study population. Although the heritability estimates varied across food groups (e.g., vegetables, fruits, meat, fish, dairy products, and snacks), the results from multiple studies suggests that genetic factors play a significant role in

explaining the variability of food intake phenotypes (Fildes et al., 2014; Hasselbalch et al., 2008; Keskitalo et al., 2008; Pimpin et al., 2013). There was evidence of age effects on the heritability of food intake phenotypes. In the Gemini study, twins aged 21 months showed lower estimates of heritability in which a heritability was estimated to be 0.15 for vegetables, 0.10 for fruits, 0.17 for dairy products, and 0.04 for snacks, compared to twins aged 3 years in which a heritability was estimated to be 0.54 for vegetables, 0.53 for fruits, 0.27 for dairy products, and 0.29 for snacks (Fildes et al., 2014; Pimpin et al., 2013). In adults, the heritability was, in general, higher than 0.20 for most food groups, including vegetables, potatoes, fruits, meat, fish, dairy products, cereal, sweets, and snacks (Hasselbalch et al., 2008; Keskitalo et al., 2008; Teucher et al., 2007).

The heritability of dietary patterns has been studied less often than for other food intake phenotypes. Most studies that have examined the heritability of dietary intake patterns performed factor analysis to identify a set of dietary patterns and used factor scores as phenotype measures (Pallister et al., 2014). In children, the heritability estimates of dietary patterns vary widely ranging from 0.12 to 0.79 in twins aged 7 y from the MacArthur Longitudinal study (Faith, Rhea, Corley, & Hewitt, 2008), from 0.20 to 0.78 in twins aged 4-5 y from the Twins Early Development study (Breen, Plomin, & Wardle, 2006), and from 0.00 to 0.90 in twins aged 7-15 y enrolled in the China Birth Defects and Child Health Care Surveillance System (Li et al., 2016). Studies in adults have produced

relatively stable estimates of dietary patterns (Bree, Eaves, & Dwyer, 1999; Keskitalo et al., 2008; Teucher et al., 2007). Although each study used the different labels for similar patterns, both healthy and unhealthy dietary patterns had heritability estimates ranging from about one-third to one-half. The heritability of "healthy foods" pattern (frequent intakes of vegetables, fruits and low-fat dairy products) was approximately 0.49 for males, and 0.54 for females, and that of "high-fat foods" pattern (frequent intakes of fried foods, fast foods, salty snacks) was 0.44 for males, and 0.47 for females in twins aged 22-27 y from the Finnish Twin Registry (Keskitalo et al., 2008). In twins aged 18-79 y from the UK female twin cohort, the estimated heritability was 0.43 for "fruit and vegetable" pattern (frequent intakes of vegetable and fruits; low consumption of fried potatoes), 0.41 for "traditional English" pattern (frequent intakes of fried foods, meats, and cruciferous vegetables), and 0.48 for "high alcohol" pattern. The heritability of both "less healthful" (frequent intakes of foods high in fat and sugar) and "healthful" (frequent intakes of vegetable, fruit, rice, and dairy products) patterns were estimated to be 0.33 in twins aged ≥ 50 y from twins enrolled in a twin registry at Virginia Commonwealth University (Bree et al., 1999).

2.3.3 Genetic Association Studies

Genetic association studies test relationships between polymorphic markers and phenotypes of interest (Lunetta, 2008). In the field of genetic epidemiology, these studies can be divided roughly into two major categories: candidate gene association studies and genome-wide association studies (GWAS). Candidate gene association studies focus on a set of single nucleotide polymorphisms (SNPs) within a single gene or a candidate region. GWAS, on the other hand, is a more encompassing and less hypothesis-driven approach which involves the analysis of a dense set of SNPs across the genome to facilitate the discovery of genotype-phenotype associations. Many results reported to date for dietary phenotypes have used a candidate gene approach, focusing on genes involved in obesity susceptibility and energy homeostasis (Grimm & Steinle, 2011).

The *FTO* has been associated with risk for obesity in multiple populations, with multiple lines of evidence supporting its role in energy balance and eating behavior (Fawcett & Barroso, 2010). Several studies have been performed to interrogate the association between variation in *FTO* and dietary phenotypes. These studies have shown that genetic variants at the *FTO* locus are associated with a variety of dietary traits, including satiety, meal frequency, total caloric and fat intake, and consumption of oils, sweets, and snacks (Cecil et al., 2008; J. M.

McCaffery et al., 2017; Jeanne M. McCaffery et al., 2012; Park et al., 2013; Timpson et al., 2008; Wardle et al., 2008).

Many studies have also focused on genes involved in the satiety signaling pathway, which includes *BDNF*, *LEP*, and *MC4R* (Walley, Asher, & Froguel, 2009). Genetic variants at the *BDNF*, which plays a prominent role in food intake regulation through central mechanisms, was associated with total caloric intake, dietary calcium intake, and snacking behavior (Dušátková et al., 2015; J. M. McCaffery et al., 2017; Jeanne M. McCaffery et al., 2012; Robiou-du-Pont et al., 2013). In the EPIC study, allelic variations in the *LEP* were found to be associated with extreme snacking (Krom et al., 2007). In the Nurses' Health Study, a common obesity variant near *MC4R* was associated with higher intake of calories and dietary fat (Qi, Kraft, Hunter, & Hu, 2008).

Genetic variations in the taste receptor genes have also been associated with food intake phenotypes (Chamoun et al., 2018; Grimm & Steinle, 2011). In particular, polymorphisms in the *taste 2 receptor member 3* gene (*TAS2R38*) were involved in the perception of bitter taste (Kim, 2003), which may contribute to over-consumption of unhealthy foods in individuals who are genetically predisposed to avoid consumption of bitter vegetables (Chamoun et al., 2018). A significant association between a genetic variant in the *taste 1 receptor member 2* gene (*TAS1R2*) and habitual consumption of sugars was also observed in two

distinct populations: diabetes-free young adults and individuals with type 2 diabetes (Eny, Wolever, Corey, & El-Sohemy, 2010).

2.3.4 Whole-genome Regression

Although genome-wide association studies have identified many loci associated with dietary intake phenotypes, the functional contributions of individual genes have not been clearly elucidated and/or explain only a small portion of the variability in dietary phenotypes. The "unexplained" heritability not accounted for in GWAS to date may be represented by thousands of unidentified variants whose effect sizes are too small to reach a genome-wide significance level (Gibson, 2010). To address this issue, whole-genome regression models have been applied to estimate the joint contribution of SNPs across all effect sizes rather than testing the association of any particular SNP to the trait of interest (J. Yang, Lee, Goddard, & Visscher, 2011). The basic concept behind this approach is to fit all the SNPs as random effects in a mixed linear model

$$\mathrm{y} = \mathbf{W}\mathbf{u} + \mathbf{e}$$

where y is a n-vector of phenotypes with n being the sample size, $\mathbf{W}$ is the standardized $n \times M$ SNP genotype matrix with M being the number of SNPs, $\mathbf{u}$ is a vector of SNPs effects with $\mathbf{u} \sim N(0, \mathbf{I}\sigma_{\mathrm{u}}^2)$ and $\mathbf{e}$ is a vector of residual effects

with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, with var(y) = $\mathbf{V} = \mathbf{W}\mathbf{W}'\sigma_u^2 + \mathbf{I}\sigma_e^2$. If we define $\mathbf{A} = \mathbf{W}\mathbf{W}'/M$ and $\sigma_G^2 = M\sigma_u^2$ as the variance explained by all the SNPs

$$y = \mathbf{g} + \mathbf{e} \text{ with } \mathbf{V} = \mathbf{A}\sigma_G^2 + \mathbf{I}\sigma_e^2$$

where $\mathbf{g} \sim N(0, \mathbf{A}\sigma_G^2)$ and **A** is a SNP-derived genetic (genomic) relationship matrix, which is estimated from genome-wide SNP data**.** Typically, whole-genome regression models use a sample of unrelated individuals. σ_G^2 is interpreted as the estimated genetic variance by contrasting the phenotypic similarity between samples to their genetic similarity. SNP-based heritability is defined as $h_{SNP}^2 = \sigma_G^2/(\sigma_G^2 + \sigma_e^2)$ and represents the proportion of phenotypic variance explained by a set of SNPs considered in the model (J. Yang et al., 2017).

Indeed, several studies have shown that analyses on genome-wide SNP data can account for a large proportion of heritability for quantitative traits, suggesting that most heritability is hiding rather than missing (Gibson, 2010; J. Yang et al., 2010; J. Yang, Manolio, et al., 2011). For example, Yang et al. (J. Yang, Manolio, et al., 2011) reported the heritability estimates of height, BMI, von Willebrand factor (vWF) and QT Interval (QTi) using 585,898 SNPs genotypes on 11,586 unrelated individuals. The joint contribution of these SNPs accounted for 45% of phenotypic variance in height, 17% of phenotypic variance in BMI, 25% of phenotypic variance in vWF, and 21% of phenotypic variance in QTi. In contrast, the set of individual causal variants identified from GWAS

cumulatively explained 10% of the phenotypic variance in height, 1.5% of phenotypic variance in BMI, 13% of phenotypic variance in vWF, and 7% of phenotypic variance in QTi.

Since whole-genome regression models use all available genetic markers jointly, their use opened new opportunities for improving the prediction of genetic predisposition to health-related traits (de los Campos, Gianola, & Allison, 2010). However, whole-gnome regression models have not been used to evaluate the contribution of genetic variation to the variability of dietary intake phenotypes in humans. In chapter 5, SNP-based heritability of dietary patterns will be estimated using whole-genome regression models with the genome-wide SNP data.

Chapter 3. Identifying Dietary Patterns among Young Adults using Sparse Latent Factor Models

ABSTRACT

Background/Objectives: Principal component analysis (PCA) has been the most widely used method in deriving dietary patterns. However, PCA requires arbitrary *ad hoc* decisions for selecting food variables in interpreting dietary patterns and does not easily accommodate other covariates. Sparse latent factor models can be utilized to address these issues. The aim of this study was to compare Bayesian sparse latent factor models to PCA for identifying dietary patterns among young adults.

Methods: Habitual food intakes were estimated using a food frequency questionnaire, generating the frequency intakes of 102 food items among 2,730 sedentary young adults aged 18-35 y (mean ± SD BMI: 26.5 ± 6.1) who exercised less than 30-min/wk for the previous 30 days without restricting caloric intake before study enrollment. Sparse latent factor modeling was applied to the standardized food intakes to derive dietary patterns, incorporating additional covariates (gender, race/ethnicity, and BMI). The identified dietary patterns via sparse latent factor modeling were compared to the PCA-driven dietary patterns.

Results: Seven dietary patterns were identified in both PCA and the sparse latent factor analysis. In contrast to PCA, the sparse latent factor analysis allowed the

covariate information to be jointly accounted for in the estimation of dietary patterns in the model and offered probabilistic criterion to determine the food relevant to each dietary pattern. The derived patterns from both methods generally described common dietary behaviors, but overall, the sparse latent factor analysis produced more distinct dietary patterns.

Conclusion: Sparse latent factor models can be useful in future studies of dietary patterns by reducing the intrinsic arbitrariness involving the choice of food variables in interpreting dietary patterns, and incorporating covariates in the assessment of dietary patterns.

3.1 INTRODUCTION

Dietary pattern studies have been undertaken in nutritional epidemiology to examine potential cumulative and interactive effects of individual components of an overall diet, in which foods are consumed in combination (Hu, 2002). In previous research, dietary patterns have been inferred theoretically based on qualitative assessments (e.g., the Healthy Eating Index (Guenther et al., 2013)), or empirically using statistical methods to extract information about dietary patterns in data (Newby & Tucker, 2004). Principal component analysis (PCA) is the most commonly used method to empirically derive dietary patterns (Gorst-Rasmussen, Dahm, Dethlefsen, Scheike, & Overvad, 2011; Varraso et al., 2012). In a dietary pattern analysis, PCA assumes that each dietary observation is characterized by a small number of latent factors (i.e., dietary patterns). Each factor is assumed to represent a multi-faceted picture of dietary consumption and to have a natural interpretation as an unobserved dietary characteristic.

Despite its popularity in dietary pattern studies, PCA has some shortcomings. The first is that it is often difficult to interpret the derived dietary patterns because they are, in general, linear combinations of *all* food variables in the observed data (Gorst-Rasmussen et al., 2011). It is a common practice to ignore foods that are weakly associated with a given pattern to simplify the interpretation. However, setting an appropriate inclusion threshold is often based

on intuition rather than on well-defined criteria (Cadima & Jolliffe, 2001; Zou et al., 2006).

Another shortcoming is that it is difficult to extend the scope of PCA to incorporate covariates that may confound measurements of dietary intake. In practice, PCA generally characterizes dietary patterns solely based on reported dietary intake and does not account for several covariates influencing individuals' dietary practices, such as age, gender, and sociocultural factors. Lack of inclusion of these covariates may lead to improper identification of latent pattern structures, since PCA attempts to find dietary patterns that explain the total variation in the dietary data as much as possible, ignoring any covariate-based trends that might contribute to this variation. To deal with non-food factors associated with dietary intake, some studies have performed stratified analyses by gender (Balder et al., 2003; Cutler, Flood, Hannan, Slavin, & Neumark-Sztainer, 2012; Flood et al., 2008; Handa & Kreiger, 2002; Nanri et al., 2017; Schulze, Hoffmann, Kroke, & Boeing, 2001; Slattery, Boucher, Caan, Potter, & Ma, 1998; Thorpe, Milte, Crawford, & McNaughton, 2016) and less often by ethnicity (Dekker et al., 2015; Nettleton, Steffen, Ni, Liu, & Jacobs, 2008; Williams et al., 2009) or age (Cutler et al., 2012; Gallagher et al., 1993), and then searched for dietary patterns therein. This strategy may not be appropriate when the global patterns are of interest in a mixed population (Fahey, Thane, Bramwell, & Coward, 2007) and it is generally limited to account for a few categorical covariates, with a small number of levels

within each category. Even if the primary interest is in the local patterns of a subpopulation, there could be unwanted sources of variance that need to be accounted for in the assessment of dietary patterns. However, simultaneous modeling of different sources of variation using both covariates and latent patterns is not a common practice in dietary pattern analysis.

Sparse latent factor models (Carvalho et al., 2008; Lucas et al., 2006; West, 2003) can be utilized to address these methodological shortcomings of PCA. West (West, 2003) initiated the idea of sparse factor modeling based on a Bayesian modeling framework, which aims to provide parsimonious relationships between high-dimensional variables and latent factors by forcing less influential associations to have a zero association in the model. This modeling framework has mainly been used in gene expression studies, but its approach is generic and broadly applicable in other fields (Carvalho et al., 2008; Lucas et al., 2006; Seo, Goldschmidt-Clermont, & West, 2007; West, 2003). Unlike PCA, sparse latent factor models naturally produce latent factors with only a subset of foods being included. Sparsity modeling is appealing in dietary pattern analysis because each pattern is typically described by a small subset of food variables in data. Taking a Bayesian approach provides a quantitative probabilistic approach for identifying the members of each subset rather than *ad hoc* threshold-based decisions that may obscure some relationships in the data. Additionally, sparse latent factor models provide a way to address potential influences of covariates. Lucas et al. (Lucas et

al., 2006) extended a sparse latent factor model by coupling it with a sparse regression model framework, where observed responses were expressed as a linear superposition of latent factors and other regressors, based on covariates. This approach allows for jointly identifying the contribution that is due to each source of variance, taking other components into account.

In this paper, we applied a sparse latent factor model to dietary pattern analysis using data from the Training Intervention and Genetics of Exercise Response (TIGER) study. Specific characteristics of participants, including gender, race/ethnicity, and BMI, were hypothesized to account for a portion of the variation observed in dietary data and jointly modeled to isolate their effects from the assessment of underlying patterns. The derived dietary patterns from sparse latent factor analysis were compared with those obtained from the classical PCA procedure.

3.2 METHODS

3.2.1 Study population

The TIGER study is a prospective cohort study, with the primary goal of identifying genetic factors that influence physiological responses to a 15-week, 3 d/wk aerobic exercise training protocol. Participant enrollment in the TIGER study was initiated in 2003 and completed in 2015. Participants were drawn from the University of Houston (2003-2008) and the University of Alabama at

Birmingham (2010-2015) in 15 cohorts ascertained each Fall or Spring semester. The target participant from this study was a sedentary individual aged 18-35 y (mean ± SD BMI: 26.5 ± 6.1) who exercised less than 30-min/wk for the previous 30 days without restricting caloric intake before enrollment. The exclusion criteria for enrollment included a physical contra-indication to exercise (e.g., cardiomyopathy), a metabolic disorder known to alter body composition (e.g., lipodystrophy), and/or pregnancy. The TIGER study was approved by the respective Institutional Review Boards, and informed consent was obtained from all participants. Details on the study protocol have been described elsewhere (Sailors et al., 2010). The present study focused on the subset of participants (n = 2,828) who reported their average intake of different foods and beverage items using a self-administered Block Food Frequency Questionnaire (FFQ, NutritionQuest, Berkeley, CA) at the baseline assessment. The FFQ was used to characterize habitual intakes of 102 food items of study participants. Food intake was measured using nine frequency categories, ranging from "never" to "every day", which were converted to frequency/day scale for each item. Participants were excluded from the analyses if they did not complete at least 90% of the food item questions, and/or were missing anthropometric or demographic data. After all exclusions, 2,730 participants were available for analysis. Participants included 1,769 female and 961 males from diverse racial/ethnic groups, including non-Hispanic white (n = 1150), African-American (n = 788), Hispanic (n = 382),

Asian (n = 166), Asian Indian (n = 81), Native American (n = 12), and others (n = 151).

3.2.2 Bayesian formulation of sparse latent factor models

Basic model form. Latent factor models aim to explain observed data in terms of a linear superposition of latent factors. Suppose that the number of observed food items is p, and the number of latent factor is k, in which $k \ll p$. For individual i, the observed food intake x_i is then expressed as

$$x_i = A\lambda_i + \nu_i$$

where A is the factor loading matrix, λ_i is latent factor scores, and ν_i is the residual noise term. Within a dietary pattern context, the factor loading matrix expresses how each dietary pattern is associated with the original food intake, factor scores represent an individual's relative position on each dietary pattern. Non-probabilistic models such as PCA do not explicitly model the noise term, but in a Bayesian context, it is usually assumed to be normal. Explicitly modeling the confounding of noise allows us to decouple variation that is unique to each food from the structure of the dietary patterns.

In order to learn the model parameters in a Bayesian context, we place prior distributions over those parameters and update these distributions with the observed data, following Bayes' law. The resulting posterior distribution then appropriately captures our beliefs and uncertainties about the parameters, and can

be used to make predictions. The model properties follow from the choice of prior distribution.

Sparse factor loadings. When modeling dietary patterns, we typically want our factor loadings to be sparse, to facilitate interpretation – if the factor loadings are sparse, each factor or dietary pattern contains only a relatively small number of significant foods. A Bayesian approach for imposing explicit sparseness on factor loadings is to assume that each loading may be zero or have non-zero with a prior probability (i.e., the probability of a factor loading being non-zero) (Carvalho et al., 2008; Lucas et al., 2006; West, 2003). More specifically, our prior distributions over individual probabilities are chosen to have substantial probability mass at zero to induce shrinkage of negligible loadings to zero. However, they also ensure that probability mass is spread over a wide range of plausible values so that important loadings escape shrinkage and take non-zero values. In our case, if a food is associated with a specific dietary pattern, the non-zero values of factor loading are modeled as arising from a normal distribution. If such is not the case, the loading will be exact zero. This reflects the fundamental view behind dietary pattern analysis that each food is expected to be largely associated with one, or at most, two pattern(s), but unrelated with the other patterns. From a modeling perspective, the loading matrix will have many zeros since each row has very few salient loadings. PCA achieves this structure in a *post hoc* manner with the use of various factor rotation techniques and loading

truncations in which the factor loadings with absolute values smaller than an arbitrary threshold are ignored. Sparse latent factor models, on the other hand, exploits sparsity-inducing priors as an integral part of the identification of dietary patterns, as described above. Based on this specification and our data, we can calculate the posterior distribution over the inclusion probabilities that are central to defining a set of foods showing significant association with specific dietary patterns. More details on sparsity-inducing priors are available in Appendix A.

Non-normality in food intake and flexible factor scores. To specify our model, we must place a prior distribution on the factor scores. This prior distribution needs to account for the arbitrary non-normal structure in dietary data that may arise from strong positive skewness and/or large proportions of zeros for many food items. A Dirichlet Process mixture model was used for the modeling of the factor score distributions in order to respond to observed non-normal structure in data (Carvalho et al., 2008). A Dirichlet Process mixture model can be thought of as a mixture of infinitely many normal distributions, and can capture any continuous distribution. The number of component distributions used is finite and determined by the data, allowing for flexible adaptation to the arbitrary non-normal structure without overfitting. It also has the feature of cutting back to normality if the data suggests that is appropriate. Full details of this approach are available in (Carvalho et al., 2008).

Incorporation of covariates. To incorporate covariate information, sparse latent factor models can be extended as described in (Carvalho et al., 2008; Lucas et al., 2006). That is,

$$x_i = \mathrm{B}h_i + A\lambda_i + \nu_i$$

where h_i is a vector of known regressors, including an intercept and B is the matrix of regression coefficients. The prior structure on regression coefficients has the same form for the factor loadings that imposes sparsity on food-covariate associations.

Inference and missing value imputations. Direct analysis of the posterior distributions over the factor loading matrix and factor scores is not feasible. Instead, we use Markov chain Monte Carlo (MCMC) methods. MCMC is a class of algorithms that generate samples from the posterior distribution over parameters. We can use these samples to make inference based on posterior means and characterize the uncertainty of model parameters. The full details of MCMC analysis for the model described in this paper are available in (Carvalho et al., 2008).

The Bayesian framework described in this paper provides a model-based solution for dealing with missing data. If the intake of some foods is missing, the missing values are treated as additional unknown parameters to be estimated. We can either integrate these parameters out, or explicitly sample them as part of our MCMC algorithm.

3.2.3 Statistical analysis

Both PCA and the sparse latent factor model were applied to the standardized intakes of 102 food items. For PCA, the missing values (0.6% of the total observations) were replaced with the median of each food item in the remaining observations. The number of components to retain was chosen based on Velicer's (Velicer, 1976) minimum average partial (MAP) test, yielding 7 dietary patterns. Varimax rotation (Kaiser, 1958) was applied to the component loadings to increase interpretability. Food items with an absolute value of loadings above 0.3 on a factor were considered a meaningful association to describe dietary patterns. PCA was performed using R, version 3.3.0 (R Core Team, 2017).

For the sparse latent factor analysis, the number of factors to retain was chosen by restricting any pattern to have at least 3 foods showing a significant association in order to make the patterns interpretable. This yielded 7 dietary patterns. The factor loading matrix was constrained to be a lower triangular matrix with positive diagonal elements to uniquely determine its structure (Aguilar & West, 2000; Carvalho et al., 2008; John Geweke & Zhou, 1996; Lucas et al., 2006). Gender, race/ethnicity, and BMI were jointly modeled to account for their contributions to the dietary measures. Cohort information was included in the model to control for subtle confounding from data collection effects. The

BFRM software (Q. Wang, Carvalho, Lucas, & West, 2007) was used to fit the sparse latent factor model.

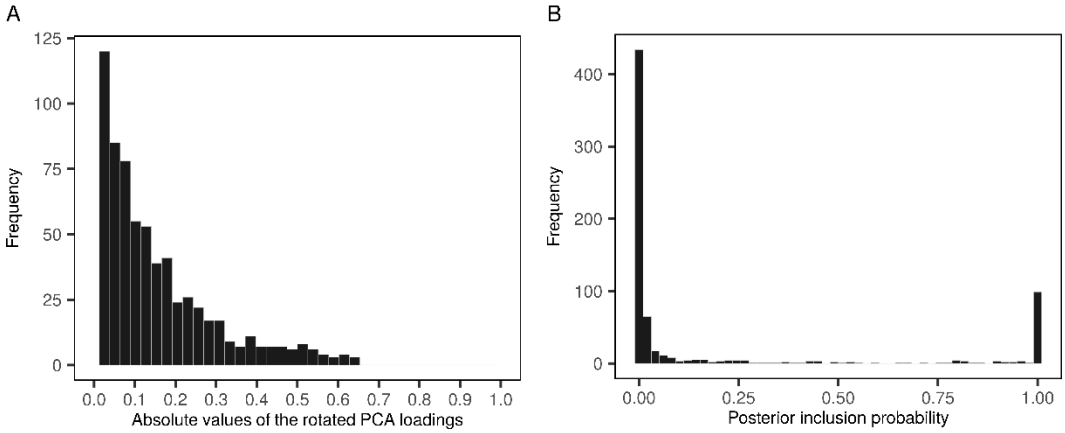


Figure 3.1: The standard interpretation of dietary patterns using PCA procedure interpret is performed by setting arbitrary thresholds, whereas sparse latent factor model uses the posterior inclusion probability that a food and a specific dietary pattern has non-zero association. (A) Histogram of absolute values of the rotated PCA loadings obtained from PCA with Varimax rotation. (B) Histogram of posterior inclusion probabilities obtained from the sparse latent factor analysis. The stronger contrast observed in (B) facilitates identifying meaningful food-pattern associations by separating them from negligible associations, compared to (A).

3.3 RESULTS

Figure 3.1A displays the absolute values of the rotated loadings from PCA with Varimax rotation that utilizes the loading-based selection of food subsets for dietary patterns. These values ranged from 0 to 0.7 without clear separation, providing little information for distinguishing which food items meaningfully

contributed to interpreting dietary patterns. Figure 3.1B displays the posterior inclusion probability that a given food is included in a specific dietary pattern. The bimodal shape of the histogram provided a strong basis for interpreting dietary patterns by screening out insignificant food-pattern pairs while highlighting those having meaningful associations.

Figure 3.2 provides a visual summary of the underlying latent factors (dietary patterns) that were ordered to increase comparability between PCA (Figure 3.2A) and the sparse latent factor analysis (Figure 3.2B). The posterior inclusion probability and the posterior mean of factor loadings with their uncertainty intervals, are available in Appendix B, which provides an image of the manner in which food items are distinctly clustered into dietary patterns. In Figure 3.2B, we show foods that have a posterior inclusion probability > 0.95 (i.e., that are relevant to a given dietary pattern with high probabilities). It should be noted that the resulting loadings between PCA and sparse latent factor model are not directly comparable since their scales are different. PCA loadings are scaled to lie between -1 and 1, but this is not the case in our approach in which each factor has its own scale that is learned from the data. In both approaches, Factor 1 displayed meaningful associations with fruit, nuts, and several vegetables. Additional foods such as fish and poultry were contained in the sparse latent factor solution, but overall, both identified a set of foods considered generally healthy. In the sparse factor analysis, Factor 2 pertained to red and processed meat, potatoes, fried

foods, pasta dishes, and breads, and Factor 3 was associated with a cluster of snacks and sweets. The PCA factors 2 and 3 also provided similar food subsets but had smaller sets of foods with more cross-loading elements between the two factors. The structure of Factor 4 was comparable between the sparse latent factor analysis and PCA, as both factors included a group of Hispanic foods, but the sparse latent factor solution formed a more concise and uncluttered cluster of foods. In PCA, it was not clear how to interpret the remaining factors. For example, Factor 5 is comprised of proteins, Factor 6 consists of beverages, and Factor 7 is a composite of somewhat unrelated foods. None of these factors would be considered as patterns of eating behavior. Conversely, the sparse latent factor model provided compact clusters of foods with clear interpretations, characterizing meat and dietary alternatives, alcoholic beverages, and cereal with milk, respectively. In addition, slight changes in the choice of cut-off had a non-negligible influence on the final pattern structures (Figure 3.3). A small decrease in the cut-off produced a more complex solution, creating many new cross-loading elements. A small increase in the cut-off provided seven clearly distinct blocks, but many foods were unrelated with any pattern. The qualitative comparison of the food subsets between two approaches is shown in Figure 3.4.

Figure 3.5, which is presented in the identical order to Figure 3.2B, displays groups of food subsets significantly associated with each of covariates. The food-covariate associations were simultaneously identified through the sparse

latent factor analysis, providing insight into the covariate-specific trends in food intake. Males consumed eggs, meat and meat products more frequently and ate vegetables, salad, yogurt, coffee, and chocolate less frequently than females. Race/ethnicity also appeared to contribute to the variation in food consumptions. African Americans consumed meat and processed foods more frequently compared to non-Hispanic whites, while consuming low-fat dairy products, fruit, and vegetables less frequently. As anticipated, Hispanics tended to eat greater amounts of traditional Hispanic foods (e.g., tortillas, beans, and tacos), and rice, compared to non-Hispanic whites. In Asian and Asian Indian groups, rice was the most notable food that distinguished both groups from the non-Hispanic white group. Asian participants consumed seafood and some vegetables more frequently than non-Hispanic whites, but such contrast was not apparent between Asian Indians and non-Hispanic whites. BMI showed little association with the frequencies of food intake.

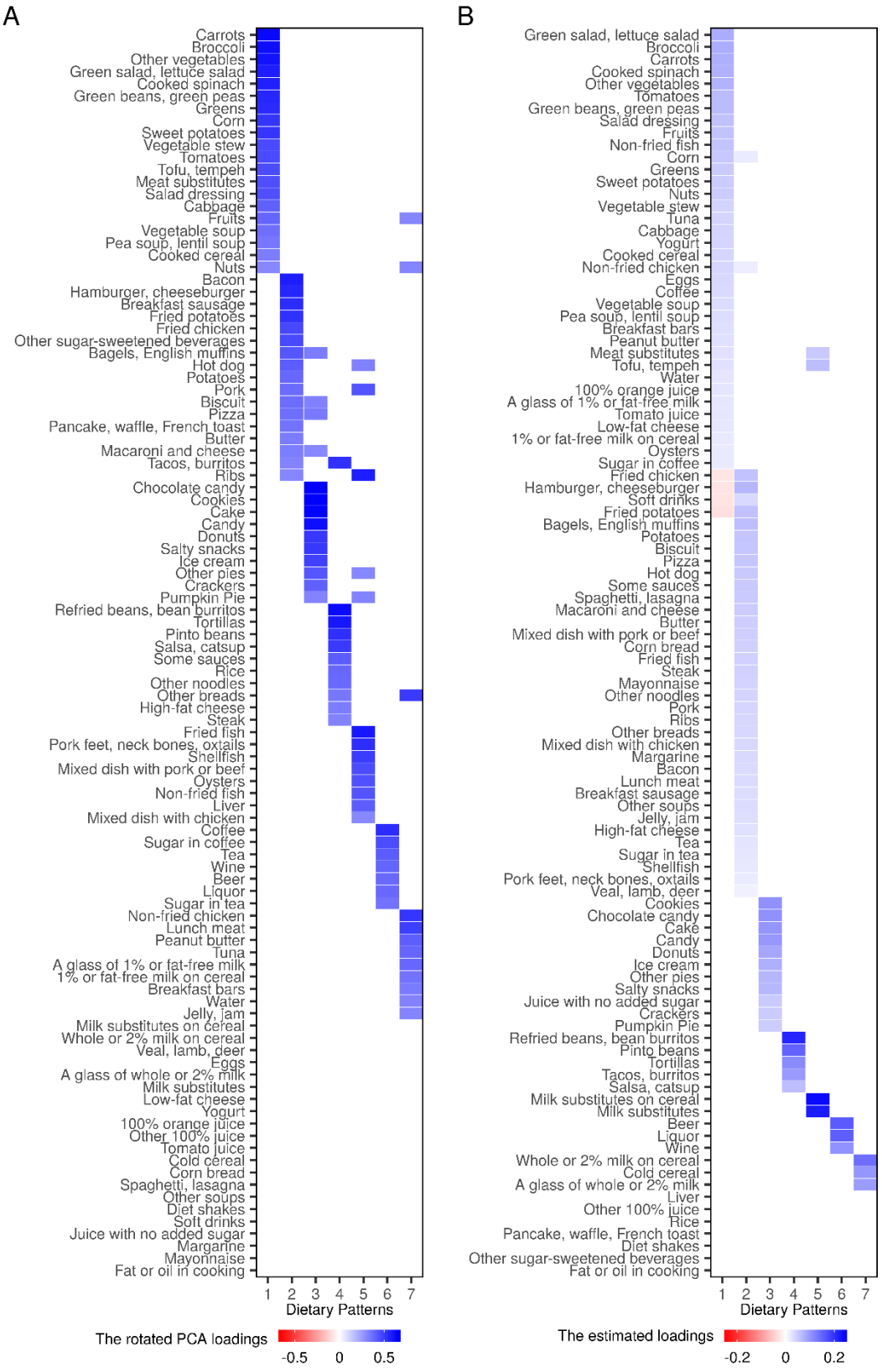

A
Carrots
Broccoli
Other vegetables
Green salad, lettuce salad
Cooked spinach
Green beans, green peas
Greens
Corn
Sweet potatoes
Vegetable stew
Tomatoes
Tofu, tempeh
Meat substitutes
Salad dressing
Cabbage
Fruits
Vegetable soup
Pea soup, lentil soup
Cooked cereal
Nuts
Bacon
Hamburger, cheeseburger
Breakfast sausage
Fried potatoes
Fried chicken
Other sugar-sweetened beverages
Bagels, English muffins
Hot dog
Potatoes
Pork
Biscuit
Pizza
Pancake, waffle, French toast
Butter
Macaroni and cheese
Tacos, burritos
Ribs
Chocolate candy
Cookies
Cake
Candy
Donuts
Salty snacks
Ice cream
Other pies
Crackers
Pumpkin Pie
Refried beans, bean burritos
Tortillas
Pinto beans
Salsa, catsup
Some sauces
Rice
Other noodles
Other breads
High-fat cheese
Steak
Fried fish
Pork feet, neck bones, oxtails
Shellfish
Mixed dish with pork or beef
Oysters
Non-fried fish
Liver
Mixed dish with chicken
Coffee
Sugar in coffee
Tea
Wine
Beer
Liquor
Sugar in tea
Non-fried chicken
Lunch meat
Peanut butter
Tuna
A glass of 1% or fat-free milk
1% or fat-free milk on cereal
Breakfast bars
Water
Jelly, jam
Milk substitutes on cereal
Whole or 2% milk on cereal
Veal, lamb, deer
Eggs
A glass of whole or 2% milk
Milk substitutes
Low-fat cheese
Yogurt
100% orange juice
Other 100% juice
Tomato juice
Cold cereal
Corn bread
Spaghetti, lasagna
Other soups
Diet shakes
Soft drinks
Juice with no added sugar
Margarine
Mayonnaise
Fat or oil in cooking
1 2 3 4 5 6 7
Dietary Patterns
The rotated PCA loadings
-0.5 0 0.5
B
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
1 2 3 4 5 6 7
Dietary Patterns
The estimated loadings
-0.2 0 0.2

Figure 3.2

Figure 3.2: The structure of latent factors is visualized using the heatmap where row correspond to food items and columns correspond to dietary patterns. Dietary patterns were identified based on the loadings for each food item using the dietary data obtained from young adults (n = 2,730) in the TIGER study. The food items were ordered to clarify the structure of dietary patterns. (A) The dietary patterns derived by PCA with Varimax rotation. The Foods with the absolute values of the rotated loading > 0.30 were used to define dietary patterns. (B) The dietary patterns derived by sparse latent factor model. Foods with the posterior mean of loadings that pass the threshold of posterior inclusion probability > 0.95 were used to define dietary patterns.

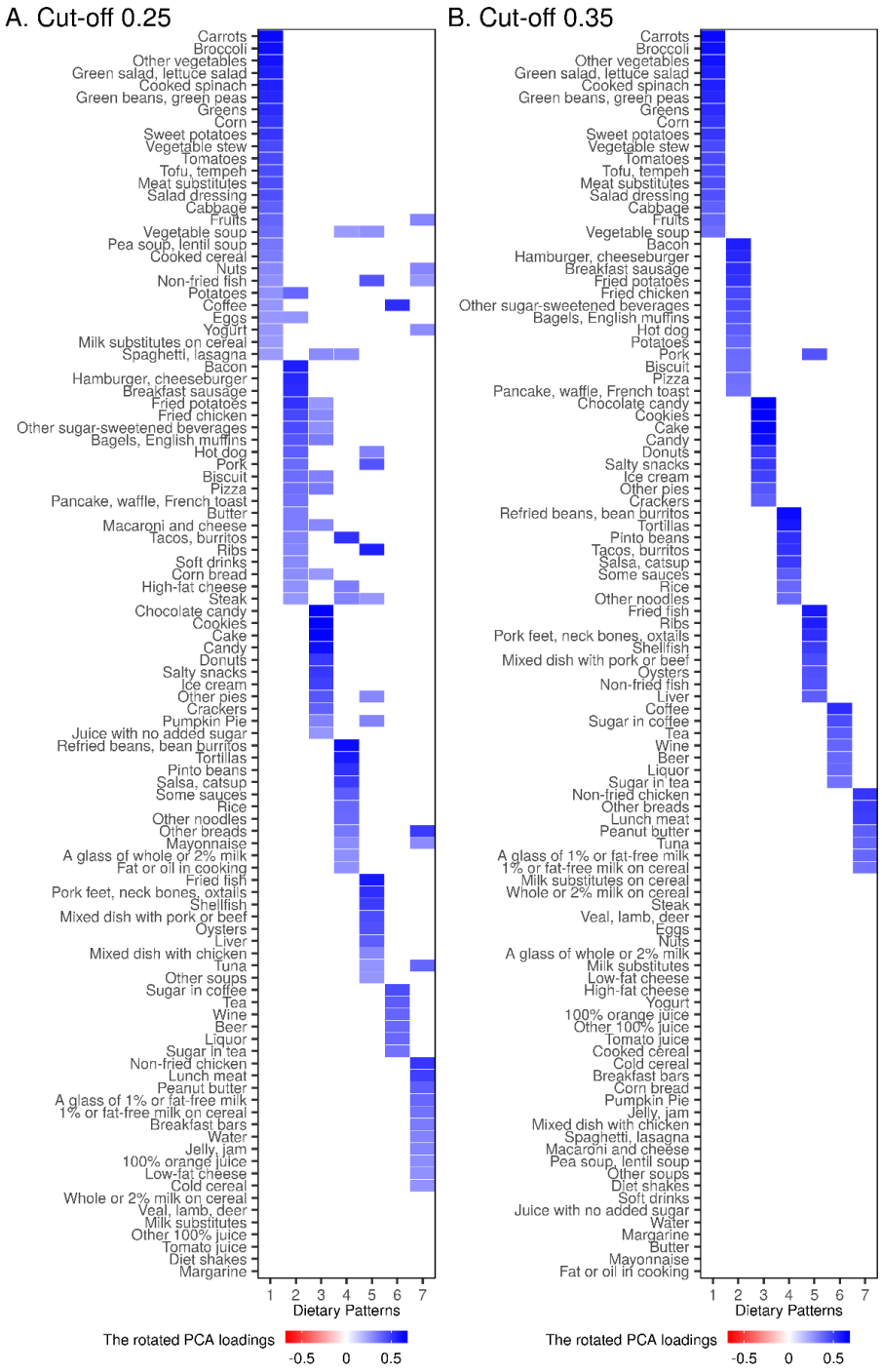


Figure 3.3: The simplified PCA-driven dietary patterns with the different values of cut-off on the magnitude of the rotated PCA loadings.

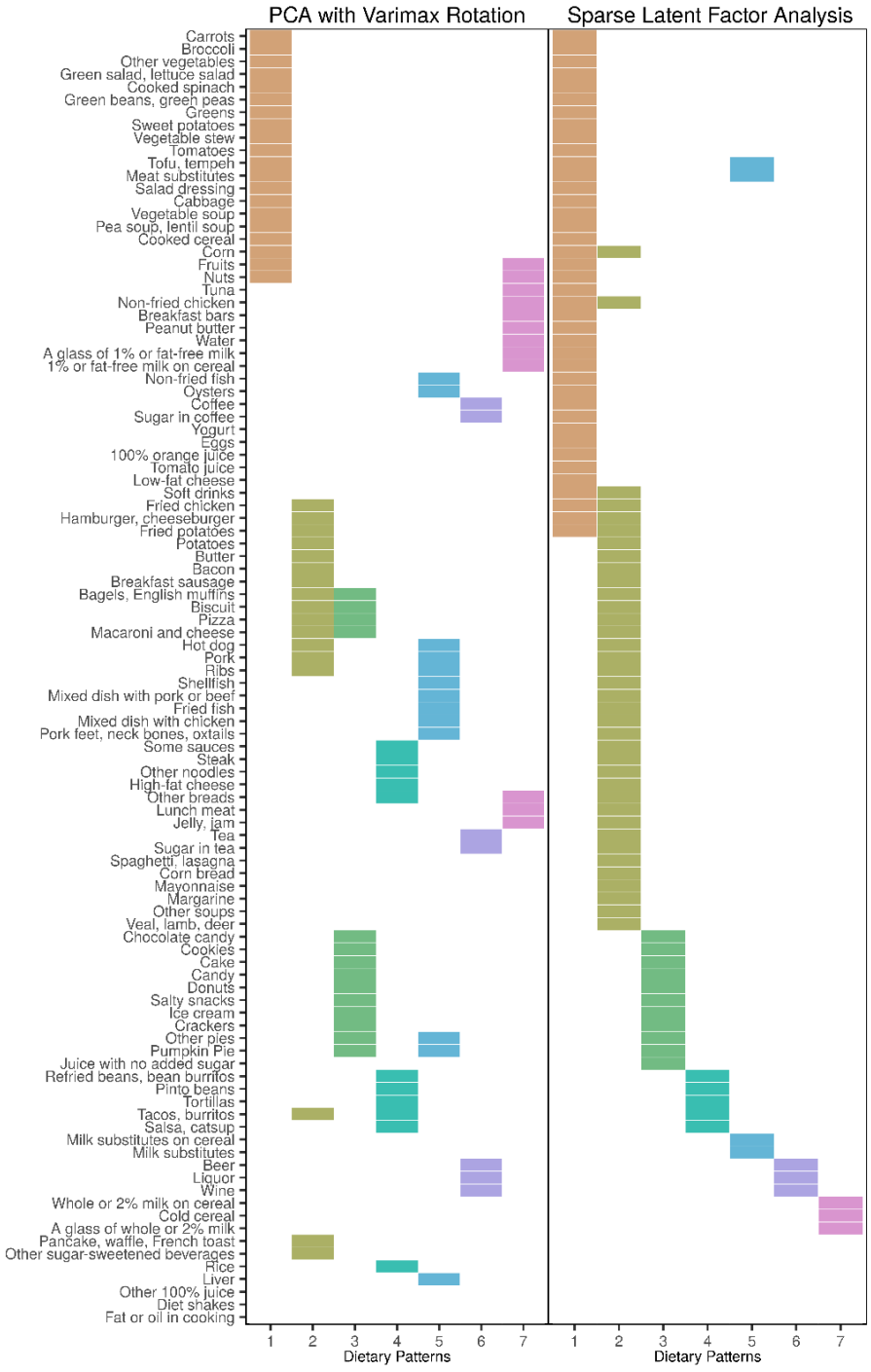


Figure 3.4: The qualitative comparison of food subsets for dietary patterns between PCA with Varimax rotation and the sparse latent factor analysis.

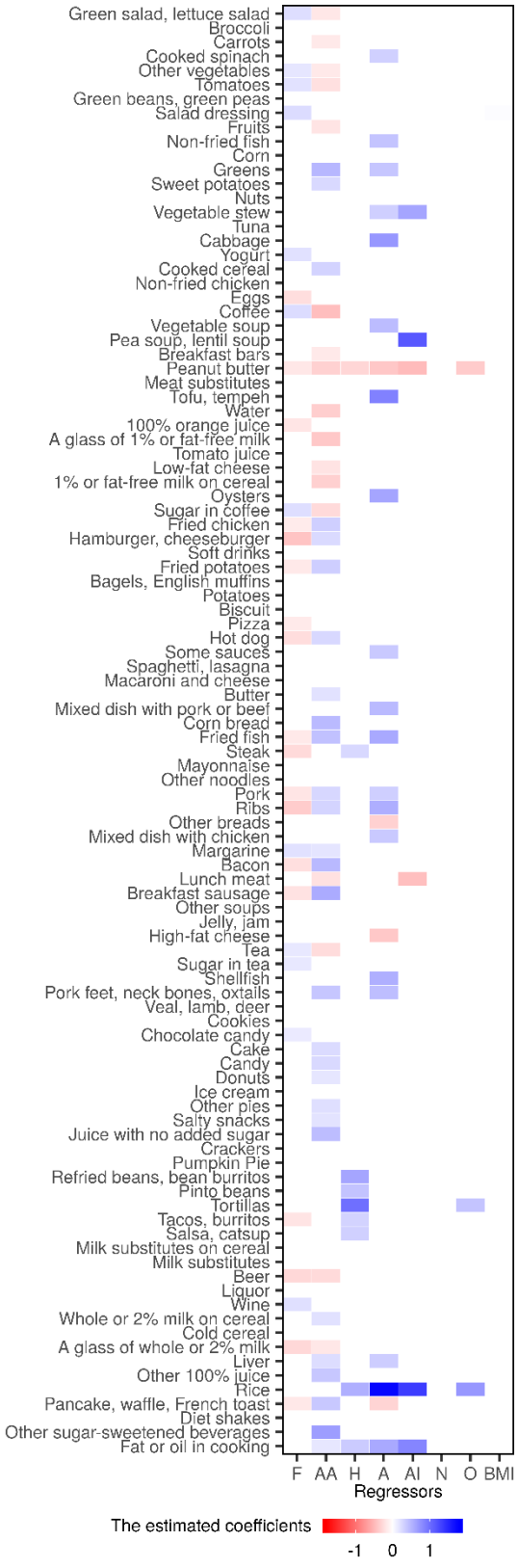

Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
F AA H A AI N O BMI
Regressors
The estimated coefficients
-1 0 1

Figure 3.5

Figure 3.5: The influences of gender, race/ethnicity, and BMI were simultaneously quantified via the sparse latent factor analysis. Gender, and race/ethnicity were dummy coded in which each of reference groups is male participants for gender category and non-Hispanic white participants for race/ethnicity category, respectively. A coefficient was set to zero if the posterior inclusion probability < 0.95. Each regressor represents the followings: (F) female, (AA) African American, (H) Hispanic (A) Asian (AI) Asian Indian (N) Native American (O) Other racial/ethnic groups, (BMI) BMI.

3.4 DISCUSSION

We performed a dietary pattern analysis using a sparse latent factor model that imposes sparsity on factor loadings to produce distinct factors comprised of only a subset of foods, while allowing for the incorporation of covariates. This sparse latent factor model was implemented in a Bayesian modeling framework that uses probability theory to represent all forms of uncertainty in the model, which in turn, allows the probabilistic assessments for the choice of food variables in interpreting dietary patterns. The sparsity-inducing prior provides strong shrinkage toward zero for foods showing trivial associations, effectively separating them from non-zero associations.

In PCA, sparsity on factor loading is typically enforced in a *post hoc* manner, by choosing a cut-off and ignoring the values below that point to define meaningful food-pattern associations. In the present study, we observed that this simple loading truncation approach may be misleading. Due to the observed weak contrast of the factor loadings in PCA, it was difficult to choose the optimal cut-off to differentiate between significant and negligible loadings, and small changes of the loading cut-point led to notable changes in the final pattern structures. In practice, the cut-off value for PCA is often determined by the ease of interpretation, but there are no objective criteria that explicitly define interpretability.

In general, nutritional epidemiological studies produce dietary data in addition to other types of data, such as demographic, anthropometric, or study design factors that may provide important information when exploring dietary patterns. In contrast to PCA, one important advantage of sparse latent factor models is their modularity that links several sub-models to address a more complex setting. A sparse latent factor model can be viewed as a multivariate regression model through a linear combination of latent factors in which regressors are themselves uncertain. Covariate information is directly integrated in this regression framework as additional predictors. Importantly, the shrinkage of our Bayesian analysis automatically takes care of the implicit multiple tests (Lucas et al., 2006; Scott & Berger, 2006) arising from the simultaneous inferences of many food-covariate associations.

In the present study, a comparison of the resulting outputs is rather complex due to their different ability to incorporate phenotypic information, but some of the factors are similar and comparable to what has been shown in the literature. In both results, Factor 1 of each result was similar to a prudent dietary pattern (Newby & Tucker, 2004; K. L. Tucker, 2010) while Factors 2 and 3 were related to a Western dietary pattern (Newby & Tucker, 2004; K. L. Tucker, 2010). A cluster of Hispanic foods was also identified in both outputs that may reflect their popularity among college students in the study cohorts (southern US). The sparse latent factor analysis produced the clusters of dairy alternatives, alcoholic

beverages, and cereal with milk, but the PCA output was not as clearly interpretable. PCA may have produced spurious associations that do not adequately reflect underlying dietary patterns, as it extracts latent factors by decomposing the total variation in the dietary intake data, without any treatment for covariate effects or measurement error.

In the sparse latent factor analysis, the identified contributions of covariates generally agree with the findings in literature. In the present study, females tended to make healthier dietary choices, which has been reported in several studies (Baker & Wardle, 2003; Fraser, Welch, Luben, Bingham, & Day, 2000; Liebman et al., 2003; Prättälä et al., 2007; Räty & Carlsson-Kanyama, 2010; Wardle et al., 2004). Comprehensive research on racial/ethnic disparities in dietary intake are still needed, but some studies have reported that non-Hispanic blacks had lower Healthy Eating Index scores (Y. Wang & Chen, 2011), as well as consumed fewer daily servings of fruits and vegetables, compared to non-Hispanic whites (Dubowitz et al., 2008). A higher consumption of rice in Asians compared to other racial/ethnic groups in the United States has also been reported (Batres-Marquez, Jensen, & Upton, 2009). Previous studies have yielded mixed results regarding the associations between dietary intake and BMI (Charlton et al., 2014; Field, Gillman, Rosner, Rockett, & Colditz, 2003; Sachithananthan & Gad, 2016; Togo, Osler, Sørensen, & Heitmann, 2001). Here, we observed little association between the food intake frequencies and BMI.

Dietary patterns are sometimes viewed as broader lifestyle patterns rather than simple diets (Martinez et al., 1998), and hence, influences of covariates are expected to be absorbed in latent pattern structures. Also, if categorical covariates, such as gender or race/ethnicity, are notably associated with food consumption, some researchers may prefer to derive dietary patterns separately for each level, instead of finding global patterns of the whole study population. Viewed in this light, it seems that the ability to accommodate covariate information is less useful. Nonetheless, the advantage of spare latent factor analysis is in its ability to robustly derive dietary patterns while simultaneously controlling for potential confounding covariates. A two-step covariate-adjustment approach has often been used in dietary pattern analysis in which each food consumption is regressed on the predefined covariate(s) and the residuals are then used as new input variables (Fahey et al., 2007; Moeller et al., 2007; Willett, Howe, & Kushi, 1997). Sparse latent factor models can provide a single-step approach to adjust for covariates by directly including them as additional regressors. The practical advantage of this approach is that it not only controls for influence of covariates, but utilizes its information jointly to derive dietary patterns. In addition, sparse latent factor models can be used without covariates. In this setting, these models are a more direct alternative to PCA.

In dietary pattern studies, one practical drawback of PCA is its inability to appropriately deal with missing values (Roweis, 1998). Since standard PCA

cannot be performed with missing values, some studies replaced missing values with the median of each food consumption or zeros before it is processed (Northstone, Ness, Emmett, & Rogers, 2008; Thorpe et al., 2016; Varraso et al., 2012). However, this strategy does not account for multivariate relationships in data. In contrast, the Bayesian aspect of sparse latent factor models allows considering missing values as unknown parameters that can be sampled in a MCMC simulation, preserving their potential dependence on other food variables, and observed covariates. The uncertainty in the imputation of missing items is propagated to other parts of the model and taken into account in the estimation of model parameters.

A common issue in modeling dietary data is that many dietary variables are often positively skewed (Hu et al., 1999). The log transformation is typically used to reduce the skewness in data, but the presence of zeros in dietary observations often requires adding an arbitrary constant to each food intake item (e.g., log(g/day intake + 1)). One other key feature of sparse latent factor models is its ability to adapt to arbitrary non-Gaussian structure in data via a Dirichlet process mixture model for the distributions of factor scores (Carvalho et al., 2008). The model exhibits non-normality distributions of factor scores when non-normality is evident in data, whereas it cuts back to normality if data is consistent with normality.

In practice, of key interest is the examination of associations between dietary pattern and health outcomes. In PCA, component scores are calculated sequentially based on the rotated loadings, which are used as explanatory variables in a regression model, including a health outcome of interest as a response variable. Sparse latent factor models also produce factor scores, but instead, they are simultaneously estimated with factor loadings. A health outcome can then be regressed on these factors scores. Other clinical predictors known *a priori* to be associated with a health outcome or for which they have a strong clinical rational can also be bundled with factor scores in this regression model, which is consistent with the conventional PCA-based approach. The ability to accommodate covariates in sparse latent factor models may offer an advantage in future health outcome analyses. The reason for incorporating covariates into sparse latent factor analysis is to derive dietary patterns that are independent of those covariates. For example, including gender in sparse latent factor analysis ensures that none of dietary patterns are highly correlated with gender. When these dietary patterns are used in conjunction with gender as regressors in the examination of relationships between dietary patterns and a health outcome, this health outcome analysis is less likely to experience a multicollinearity problem. The inclusion of gender in sparse latent factor analysis does not preclude the use of gender in health outcome analyses, since gender may have its own influence on health outcomes, independent of its relationships with dietary patterns. Moreover,

as a probabilistic module, these models have a potential to be further extended to include health outcome variables so that estimations of factor loadings and scores, and examination of pattern-health relationships are jointly performed in one coherent procedure. Coupling predictive regression components for different types of health outcomes (e.g., binary, categorical, and censored) to dietary pattern modeling would be an interesting area of future work.

While we demonstrated that sparse latent factor models may offer several advantages over PCA, there is still an ongoing debate as to whether the use of sparsity is appropriate in dietary pattern analysis (Gorst-Rasmussen et al., 2011; Imamura & Jacques, 2011). Imamura and Jacques (Imamura & Jacques, 2011) argue that sparsity may be less useful than other methods because the cumulative effect of all foods is important when considering outcomes and risk associated with dietary intake. Nevertheless, interpretable pattern structures are obtained via selection of food variables in practice. This is the rationale behind the use of loading truncations on the dietary patterns produced by PCA with subsequent rotation. Sparsity can be thought of as an alternative way of regularizing dietary pattern modeling in an attempt to better explain the variation structure in data, not to make a statement that associations are truly zero. The sparsity approach helps to pull out the foods with dominant effects by allowing many others to be zero if their contributions are small enough.

The current study comes with limitations. The present analysis used single food items as input variables for dietary pattern analysis, whereas food items are commonly collapsed into the smaller number of food groups in the PCA approach. Moeller et al. (Moeller et al., 2007) pointed out that using too many food variables may lead to odd combinations with undue influence. However, an investigator-driven food grouping strategy can be subject to arbitrary choices, since there are numerous ways of aggregating food items differently into different numbers of food groups (Martinez et al., 1998; *Scientific Report of the 2015 Dietary Guidelines Advisory Committee*, 2015). Indeed, the Nutrition Evidence Library review (*A Series of Systematic Reviews on the Relationship Between Dietary Patterns and Health Outcomes*, 2014) reported that there were variations in the number and type of food grouping across dietary pattern studies. Future studies designed to evaluate the performance of two methods in different input variables settings and improve standardization of food grouping scheme would greatly enrich our understanding of dietary pattern analysis. It also should be noted that the present study relied on self-reported food frequency questionnaire to assess food intakes, which are subject to measurement errors.

A potential barrier of the use of sparse latent factor models could be its mathematical complexity in comparison to PCA. It is true that PCA is relatively simple, and works in many settings with careful scrutiny. However, it is also clearly desirable to have a flexible, extensible approach that can be adapted

individually to specific study situations and formalize our intuition as well-defined criteria. In this respect, sparse latent factor models can be complementary to PCA by allowing investigators to move closer to modeling the complex situations when it is necessary.

The proper characterization of dietary patterns is important in accounting for inherent interactions among foods that may have synergistic and cumulative effects on health outcomes, describing common characteristics of eating behaviors, and identifying features belonging to a healthy diet to provide guidance for nutrition intervention. The classical PCA procedure in dietary pattern analysis does not properly handle the intrinsic arbitrariness inherent in the choice of food variables for characterizing dietary patterns, the contribution of covariates to variation in data, and other sources of uncertainty arising from noise and missing values in the measurement process. Progress can be made in understanding patterns of dietary intake if all reasonable sources of uncertainty and variation are incorporated in the evaluation of dietary patterns. The sparse latent factor models can be a useful addition to dietary pattern modeling by addressing the practical issues in PCA.

Chapter 4. The Influence of 15-week Exercise Training on Dietary Preferences among Young Adults

ABSTRACT

Background/Objectives: Little is currently known about how exercise may influence dietary patterns and/or food preferences. The present study aimed to examine the effect of 15-week exercise training on overall dietary preferences among young adults.

Methods: This study consisted of 2,680 young adults drawn from the Training Intervention and Genetics of Exercise Response (TIGER) study. Subjects underwent 15 weeks of aerobic exercise training, and exercise duration, intensity, and dose were recorded for each session using computerized heart rate monitors. In total, 4,355 dietary observations with 102 food items were collected using a self-administered food frequency questionnaire before and after exercise training (n = 2,476 at baseline; n = 1,859 at 15 week). Dietary patterns were identified using a Bayesian sparse latent factor model. Changes in dietary preferences were evaluated based on the pre/post-training differences in dietary pattern scores, accounting for the effects of gender, race/ethnicity, and BMI.

Results: Within each of the seven dietary patterns identified, most dietary pattern scores were decreased following exercise training, consistent with increased

voluntary regulation of food intake. A longer duration of exercise was associated with decreased preferences for the Western (β: -0.0793; 95% credible interval: -0.1568, -0.0017) and Snacking (β: -0.1280; 95% credible interval: -0.1877, -0.0637) patterns, while a higher intensity of exercise was linked to an increased preference for the Prudent pattern (β: 0.0623; 95% credible interval: 0.0159, 0.1111). Consequently, a higher dose of exercise was related to a decreased preference for the Snacking pattern (β: -0.0023; 95% credible interval: -0.0042, -0.0004) and an increased preference for the Prudent pattern (β: 0.0029; 95% credible interval: 0.0009, 0.0048).

Conclusion: The 15-week exercise training appeared to motivate young adults to pursue healthier dietary preferences and to regulate their food intake.

4.1 INTRODUCTION

The transition from adolescence to young adulthood is associated with a variety of major lifestyle changes that are likely to influence health-related behaviors (Gordon-Larsen, Nelson, & Popkin, 2004). Young adults have more independence over their dietary choices, while still being financially limited, which may lead to poor eating habits, and positive energy balance (D. A. Anderson, Shapiro, & Lundgren, 2003; Deliens, Clarys, De Bourdeaudhuij, & Deforche, 2014; Franko et al., 2008; Kattelmann et al., 2014; Kolodinsky, Harvey-Berino, Berlin, Johnson, & Reynolds, 2007; Plotnikoff et al., 2015; Wengreen & Moncur, 2009). Previous studies have reported that considerable weight gain takes place during the college years, and being mildly to moderately overweight at age 20 to 22 years is associated with increased risk for obesity in later years (Fedewa, Das, Evans, & Dishman, 2014; Gropper, Simmons, Connell, & Ulrich, 2012; Kattelmann et al., 2014; McTigue, Garrett, & Popkin, 2002; Plotnikoff et al., 2015). Thus, it is important to identify an effective intervention strategy targeting this life cycle transition to prevent weight gain and obesity that may be less malleable in adulthood.

Exercise contributes to both the prevention and treatment of obesity. One potential mechanism associated with exercise for improving health is through altering dietary behavior (Blundell et al., 2015; C. Martins et al., 2008). The majority of studies that have measured the influence of exercise on dietary

behaviors have been focused on energy intake in relation to appetite (Blundell et al., 2015; Donnelly et al., 2014; Elder & Roberts, 2007; King, Burley, Blundell, & others, 1994; C. Martins et al., 2008). To date, there is no definite consensus that exercise modulates appetite or energy intake (Donnelly et al., 2014). The influence of exercise on dietary preference has been studied less often than energy intake. Previous studies showed inconsistent results regarding whether exercise motivates people to alter their macronutrient intake (Donnelly et al., 2014; Elder & Roberts, 2007; Catia Martins et al., 2015; Schubert et al., 2013). However, evidence suggests that exercise may play a role in specific dietary preferences. In rodents, wheel running led to a decrease in preferences for a high fat diet that was, in part, sex dependent (Chen et al., 2017; Lee et al., 2017; Liang et al., 2015; Moody, Liang, Choi, Moran, & Liang, 2015; H.-J. Wang et al., 2017; T. Yang et al., 2017). In humans, physical exercise has been associated with dietary preferences for sweet substances, salt, high fat foods, and high energy density foods (Horio, 2004; Horio & Kawamura, 1998; Horsch et al., 2015; Kanarek et al., 1995; Leshem et al., 1999; McNeil et al., 2015; Panek et al., 2014). Considering its role in healthy weight maintenance and weight loss (J. W. Anderson, Konz, Frederich, & Wood, 2001; Swift, Johannsen, Lavie, Earnest, & Church, 2014), exercise may provide behavioral adaptation to counteract obesity, partly through changes in dietary preference. Taken together, understanding the

role of exercise in dietary preferences among young adults is important in maintaining lifelong health and designing effective health intervention programs.

Though the available evidence supports the potential role of exercise on dietary preferences, most of the research on this topic has focused on isolated categories within a diet or on single nutrients. The primary limitation of this approach is that it may fail to acknowledge the interactive effects of multiple dietary components. When an individual increases or restrains his/her intake of one dietary component, it is often accompanied by changes in other aspects of eating behavior. Given that individuals eat foods and nutrients in a variety of combinations under free-living circumstances, the impact of exercise on dietary preferences may be more complicated than previously described. Accordingly, it is necessary to examine how exercise and its characteristics influence individuals' preferences for overall eating patterns.

A data-driven dietary pattern analysis has been used in the field of nutritional epidemiology as a means of characterizing habitual eating patterns. This approach typically assumes that the collection of observed food intakes can be expressed as a small set of underlying patterns in data, in which each pattern is comprised of a multi-faceted composite of dietary consumption that has a natural interpretation as an unobserved dietary characteristic. Importantly, pattern analysis generates a vector of pattern scores that describes how much each dietary pattern contributes to observed dietary intake, allowing for the examination of the

transition of dietary preferences quantitatively by evaluating the changes in these scores.

In the present study, we examined the influence of a 15-week exercise training protocol on dietary preferences in young adults with different exercise characteristics including exercise intensity, duration, and dose. Habitual dietary patterns were derived using data from the Training Intervention and Genetics of Exercise Response (TIGER) study, and transitions of dietary preferences following exercise training were evaluated using data-driven dietary pattern analysis. We also examined how gender and race/ethnicity interact with exercise in influencing dietary preferences.

4.2 METHODS

4.2.1 Study population

The data were collected as part of the TIGER study, which is a prospective cohort study with the goal of introducing sedentary young adults to regular exercise via a 15-week, 3 d/wk aerobic exercise training provided in the form of a course for college credit. Between 2003 and 2015, sedentary college students who had exercised less than 30 min/wk without caloric restriction for the last 30 days prior to participation were recruited at the University of Houston (2003-2008) and the University of Alabama at Birmingham (2010-2015). Exclusion criteria for study participation included: a physical contraindication to exercise (e.g.,

cardiomyopathy), pregnancy, and/or a known metabolic disorder that may influence body composition. The study was approved by the Institutional Review Board of each participating institution, and written informed consent was obtained from all study participants.

4.2.2 Exercise assessments

The prescribed exercise sessions consisted of 30 min aerobic exercise at 65%-85% of age- and gender-specific maximum heart rate reserve, along with a 5 min warm-up and a 5 min cool-down. During each exercise session, participants wore computerized heart rate monitors (Polar Inc, Bethpage, NY) that recorded minute-to-minute heart rate, date, time and duration of each exercise session. Participants chose from a variety of aerobic exercise modes, including treadmill, elliptical trainer, stair stepper, or exercise bike. Participants were permitted to exercise for longer durations (up to 60 min) than the prescribed session. More details on the exercise protocol are available elsewhere (Miller et al., 2014; Sailors et al., 2010).

In this study, exercise duration was recorded in minutes, and an average exercise intensity for each session was calculated based on percent of heart rate reserve (%HRR). Total exercise dose was quantified using a heart rate physical activity score (HRPAS) that adjusts exercise duration in minutes by exercise intensity for each session, which were then summed over all workout sessions (Miller et al., 2014). Exercise compliance was evaluated by comparing observed

values of the HRPAS score to prescribed values, calculated on the basis of a minimum %HRR of 65% for at least 30 min per session, summed over all possible sessions.

4.2.3 Dietary assessment

Dietary intake was assessed via the Block Food Frequency Questionnaire (FFQ, NutritionQuest, Berkeley, CA), providing habitual intakes of 102 food items of study participants. Participants were asked to answer how often they consumed each food item using nine frequency categories, ranging from “never” to “every day” at baseline and at 15 weeks. Each food intake was converted to a daily frequency and then was standardized by subtracting its mean and dividing the difference by its standard deviation. In the present analysis, dietary observations were excluded if participants were missing demographic or anthropometric data, had invalid exercise data (e.g., technical errors with the heart rate monitors), and/or had more than 10 missing responses of FFQ food item questions. In total, 4,355 dietary observations (n = 2,476 at baseline; n = 1,859 at 15 weeks) were available, where 1,655 participants had valid dietary observations at both baseline and 15 weeks, 821 participants had an observation only at baseline, and 204 participants had an observation only at 15 weeks for a total of 2 680 individuals. This samples included 1,730 women and 950 men from diverse racial/ethnic groups, including non-Hispanic white (n = 1,145), African-American

(n = 784), Hispanic (n = 361), Asian (n = 156), Asian Indian (n = 78), and others (n = 156).

4.2.4 Statistical analysis

The average changes of anthropometric measures and exercise parameters are summarized using means and standard deviations (SD). A Bayesian sparse latent factor model was applied to the baseline dietary data in order to derive habitual dietary patterns. In this model, a factor loading matrix expresses how each dietary pattern is associated with observed food intakes, and factor score (i.e., dietary pattern score) represents an individual's relative position on a given dietary pattern. In the present study, sparsity constraints enforced via a Bayesian variable selection technique were imposed on factor loadings, providing a posterior inclusion probability for each factor loading (Carvalho et al., 2008; Lucas et al., 2006; West, 2003). This probability reflects the degree to which model favors the inclusion of a food in each of latent factors, which are central to interpreting dietary patterns. Here, food with posterior inclusion probability (i.e., food relevant to a given dietary pattern with a high probability) were used to define a food subset for each dietary pattern. The number of dietary pattern were determined by retaining all factors including at least 3 foods with posterior inclusion probability above the cut-off. The factor loading matrix were assumed to remained unchanged after 15 weeks so that the interpretation of dietary patterns does not changes over time, and factor scores at different time points are

comparable on the same basis. On the other hand, factor scores are assumed to vary at different time points, allowing the changes of dietary patterns are captured by scores differences before and after the 15-week exercise training. A multivariate linear mixed-effects model was used to estimate factor scores at baseline and at 15 weeks. In this study, the random effects component was used to account for individual-dependent and cohort-dependent variation not capture by the fixed effects component that includes linear predictor, such as gender, race/ethnicity, and exercise parameters. More details on the model specification for the estimation of factor scores are available in Appendix C.

We first evaluated the general trend of the exercise intervention effect on dietary preferences among all participants, with covariates including gender, race/ethnicity, and BMI. Gender and race/ethnicity were dummy-coded, with male and non-Hispanic white acting as the reference group of each category, respectively. Next, we investigated how the influence of exercise training on dietary preferences varies with different components of exercise characteristics, gender, and race/ethnicity by evaluating their interactions with the time variable. Model parameters were estimated in a Bayesian context by placing prior distributions over those parameters and updating these distributions with the observed data. Inference is based on the posterior distribution via samples obtained from a Markov Chain Monte Carlo method. Model estimates were summarized using their posterior means with 95% credible intervals that

represented an interval having a 95% probability of containing the true value of the model parameter. Missing values in food intakes were dealt with using a model-based imputation approach in which missing values were imputed at each iteration of the Markov Chain Monte Carlo sampler. All analyses were performed using the BFRM software (Q. Wang et al., 2007) and Stan (Carpenter et al., 2017). Figures were generated using the ggplot2 package (Wickham, 2009, p. 2) in R version 3.4.3 (R Core Team, 2017).

4.3 RESULTS

Table 4.1 shows the descriptive statistics of exercise parameters and changes of anthropometric measurements following the exercise treatment. Among 2,680 participants, 2,053 participants were identified as exercise compliant, while 627 participants were classified as noncompliant on the basis of the prescribed exercise dose. Exercise compliance was associated with generally more positive changes in anthropometric measures, compared to noncompliance (Herring, Sailors, & Bray, 2014).

Table 4.1: Descriptive statistics of exercise characteristics and changes of anthropometric variables (n = 2,680).

	Mean (SD)
Exercise Measures	
Mean duration (min/session)	38.05 (4.65)
Mean intensity (%HRR)	66.70 (5.80)
Total dose (HRPAS)	774.66 (223.63)
Attendance rate (%)	83.11 (20.91)
Anthropometric Measures	
BMI change ($kg \cdot m^{-2}$)*	-0.19 (0.95)
Body Mass change (kg)**	-0.38 (2.66)
Hip change (cm)**	-0.67 (4.10)
Waist change (cm)*	-0.99 (3.84)
DXA fat change (%)**	-0.59 (1.73)

DXA: Dual-energy X-ray absorptiometry

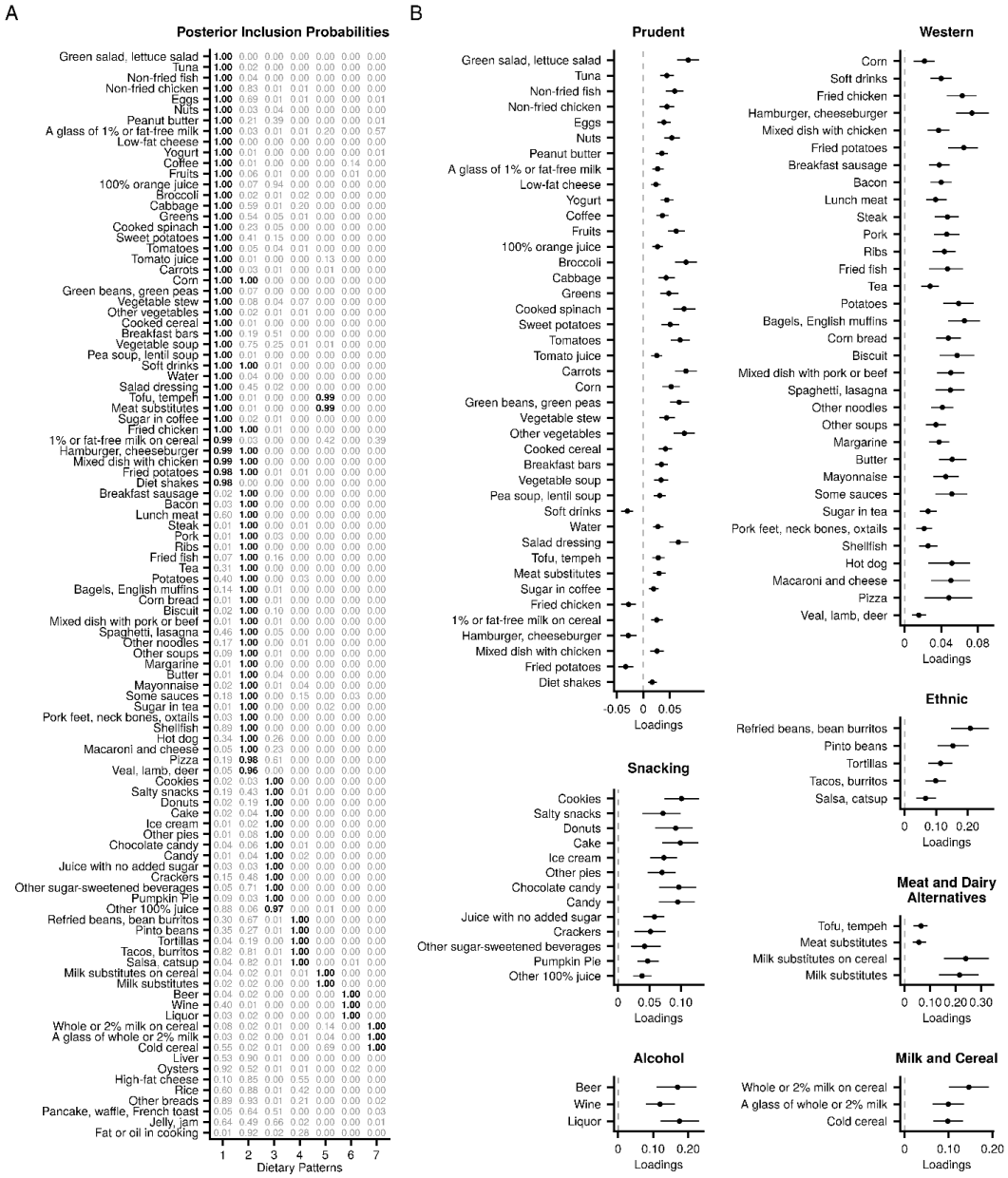


Figure 4.1: Dietary patterns among young adults identified in the TIGER study. (A) Posterior inclusion probabilities of individual food-factor pairs. Foods showing posterior inclusion probability > 0.95 (boldface) for each factor were considered in defining dietary patterns. (B) Posterior means of significant factor loadings with 95% credible intervals are displayed for each dietary pattern, respectively.

Figure 4.1 displays seven habitual dietary patterns identified from the TIGER study participants, which were labeled as Prudent, Western, Snacking, Ethnic, Meat and Dairy Alternatives, Alcohol, and Milk and Cereal. Dietary patterns were interpreted based on posterior inclusion probabilities of individual food-factor pairs (Figure 4.1A) and the factor loading estimates above the probability threshold (Figure 4.1B). The Prudent pattern was mainly associated with a high frequency intake of fruits, vegetables, and low-fat foods and a low frequency intake of fried foods, and soft drinks. The Western pattern was primarily associated with a high frequency intake of red meat, processed meat, fried foods, soft drinks, breads, and pasta dishes. The Snacking pattern was associated with a high frequency intake of foods such as cookies, sweets, and salty snacks. The Ethnic pattern showed positive associations with foods such as tacos, burritos, tortillas, and salsa. The remaining patterns were designated based on their food subsets.

Figure 4.2 displays the treatment effects of exercise training on dietary pattern scores. Most dietary patterns scores were decreased or unchanged after the exercise intervention. This trend was confirmed in the analysis performed at the individual food item level in which most of food frequency intakes were decreased at 15 weeks, as shown in Figure 4.3. There were insignificant changes in the scores of Meat and Dairy Alternatives and Alcohol patterns between before and after exercise training.

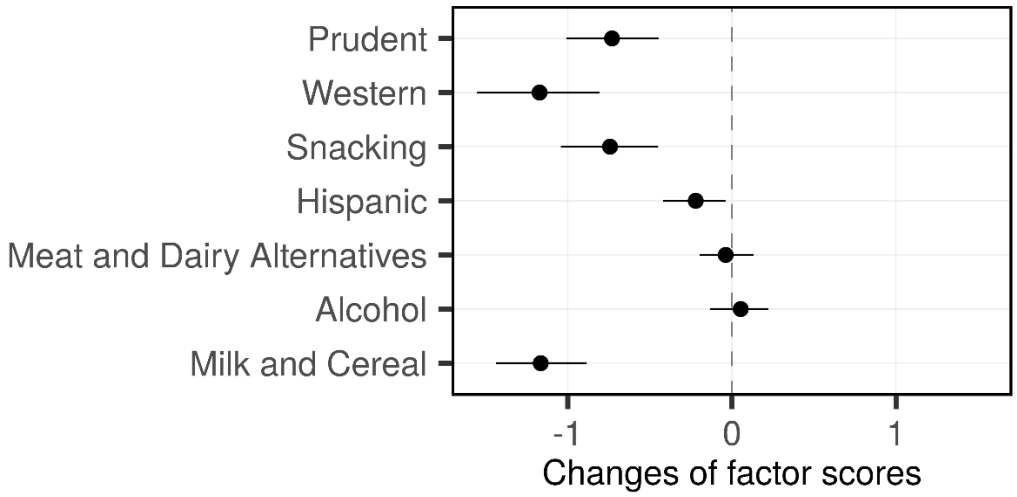


Figure 4.2: Exercise treatment effect on dietary preferences based on changes in dietary pattern scores for all participants. Changes of pattern scores were summarized using their posterior means with 95% credible intervals. Most dietary pattern scores were decreased following exercise training, except for the Meat and Dairy Alternative and Alcohol patterns.

Figure 4.4 displays how changes in dietary pattern scores vary with different values of exercise duration, intensity, and dose. A longer duration of exercise was associated with a decrease in preferences for foods in the Western (β: -0.0793; 95% credible interval: -0.1568, -0.0017), and Snacking (β: -0.1280; 95% credible interval: -0.1877, -0.0637) patterns and an increase in preference for the Milk and Cereal pattern (β: 0.0700; 95% credible interval: 0.0107, 0.1258) (Figure 3A). A higher intensity of exercise was linked with a greater preference for the Prudent pattern (β: 0.0623; 95% credible interval: 0.0159, 0.1111) (Figure

4.4B). Accordingly, participants who achieved a higher dose of exercise were more likely to increase their preference for the Prudent pattern (β: 0.0029; 95% credible interval: 0.0009, 0.0048) but reduce the preference for the Snacking pattern (β: -0.0023; 95% credible interval: -0.0042, -0.0004) (Figure 4.4C). Figure 3D displays the effects of exercise on dietary pattern preferences as a function of exercise compliance. Participants who achieved the prescribed dose of exercise (i.e., compliant), were more likely to maintain their preferences for the Prudent pattern following exercise training, compared to their non-compliant counterparts. In addition, participants who were exercise compliant showed a decrease in the preference for the Snacking pattern, whereas those who were noncompliant exhibited an increased in the Snacking pattern.

Figure 4.5 displays how the influence of exercise on dietary preference varies by gender or race/ethnicity. In this study, no gender differences were observed in the changes in dietary preferences following exercise training (Figure 4.6A). In contrast, the race/ethnicity of participants appeared to be associated with different patterns in dietary preferences. Asian participants showed a distinct trend in which the Western pattern score was increased after exercise training (β = 1.9006; 95% credible interval = 0.3093, 3.4390), in contrast to all other racial/ethnic groups which experienced a decrease in the Western pattern following exercise training (Figure 4.5B). Hispanic participants showed a decreasing trend in the Ethnic pattern score, which was not seen in other

racial/ethnic groups (β = -1.0153; 95% credible interval = -1.5565, -0.4775) (Figure 4.5C). Asian Indians showed a similar trend for the Ethnic pattern score, but with weaker evidence (β = -0.9186; 95% credible interval = -2.0636, 0.1555).

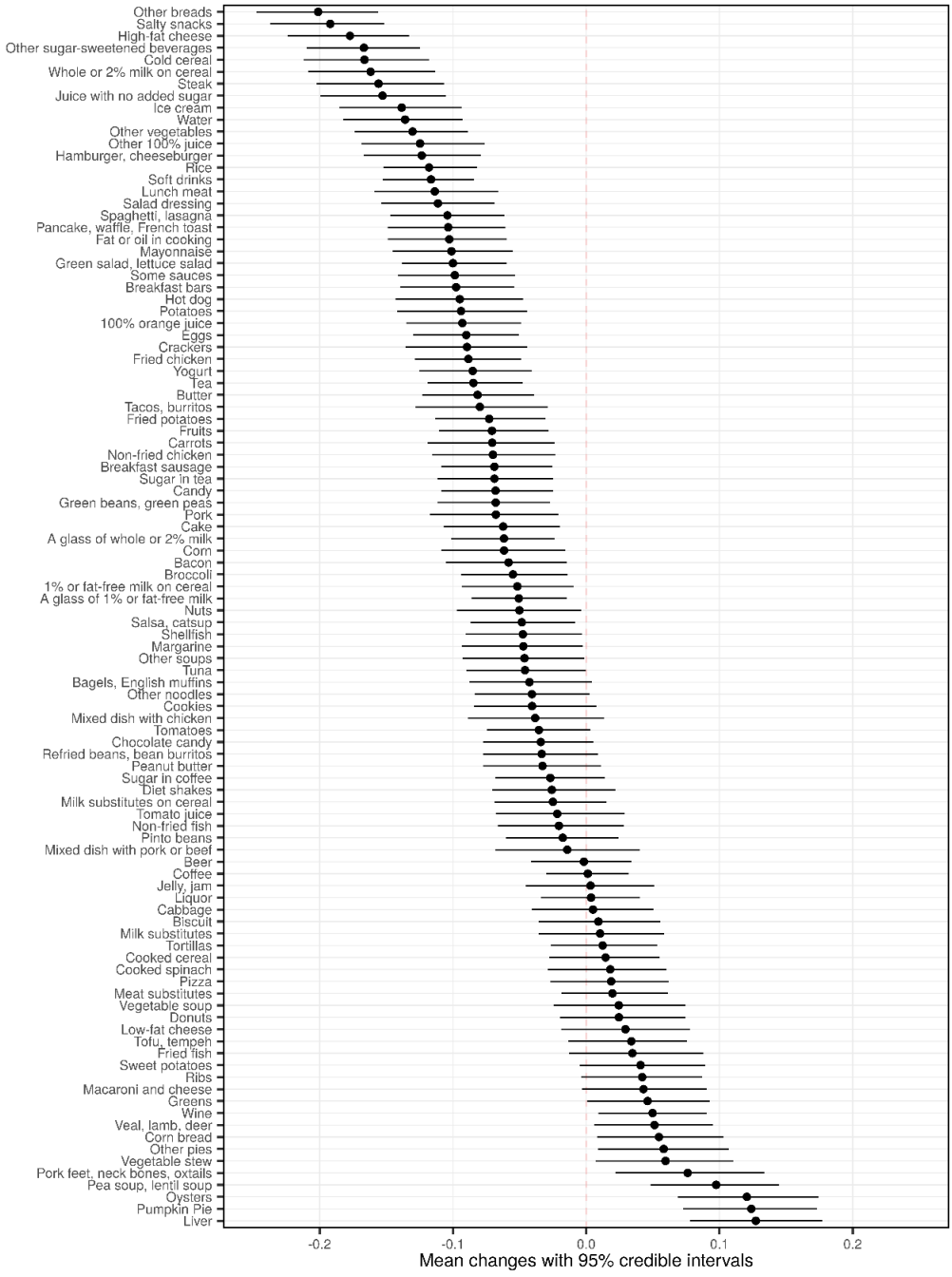


Figure 4.3: Changes of individual food item intakes based on a multivariate linear mixed model. Gender, race/ethnicity, and BMI were adjusted. The results were summarized using their posterior means with 95% credible intervals.

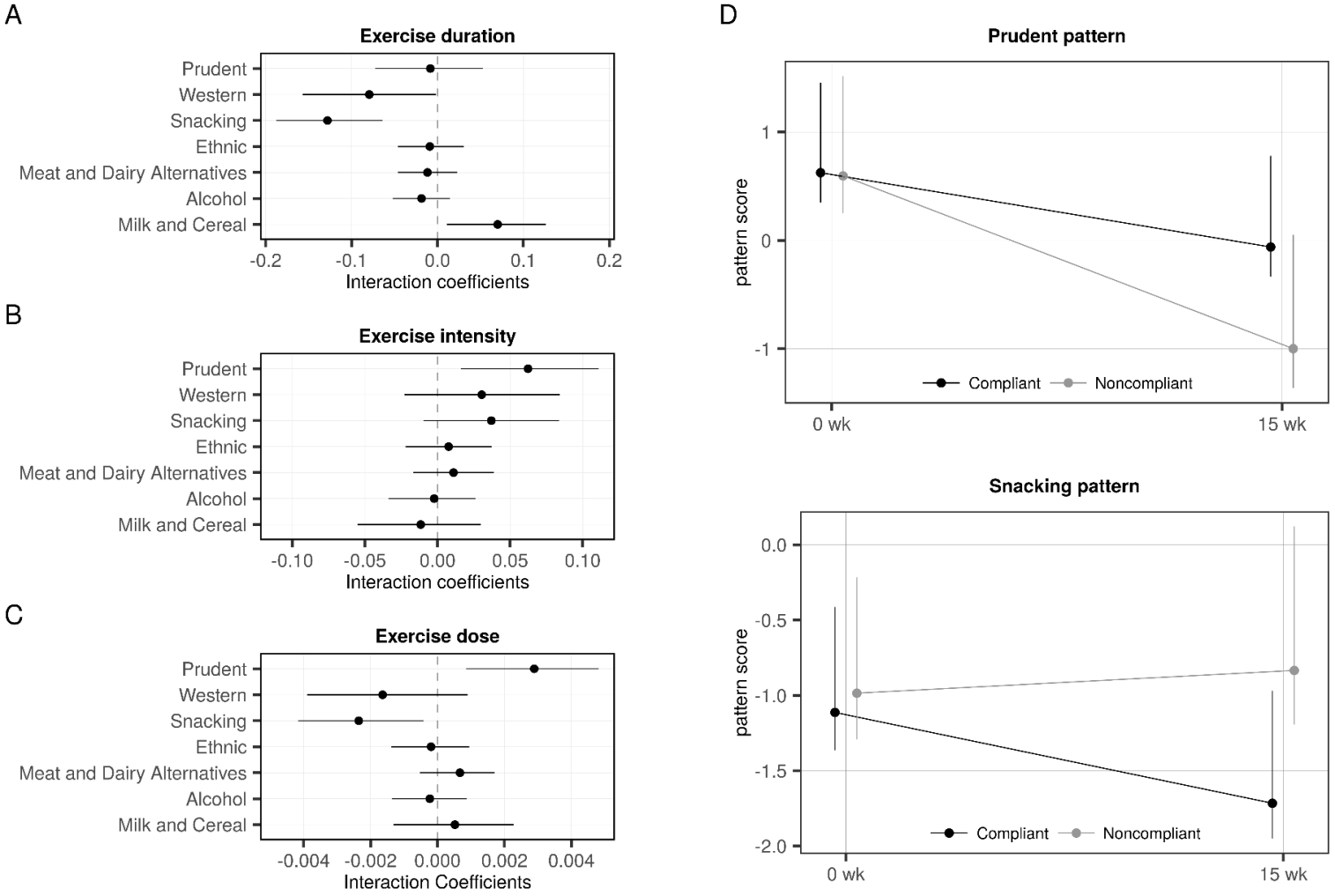


Figure 4.4: Influences of exercise characteristics on the treatment effect of exercise intervention on dietary preferences. Subfigures A-C represent how treatment effect on dietary preferences vary by each one-unit increase of exercise duration, intensity, and dose. Results were summarized with their posterior means of 95% credible intervals. (A) Influence of exercise duration (min). (B) Influence of exercise intensity (%HRR). (C) Influence of exercise dose (HRPAS) (D) Changes of the Prudent and Snacking pattern scores by a dose-based exercise compliance status.

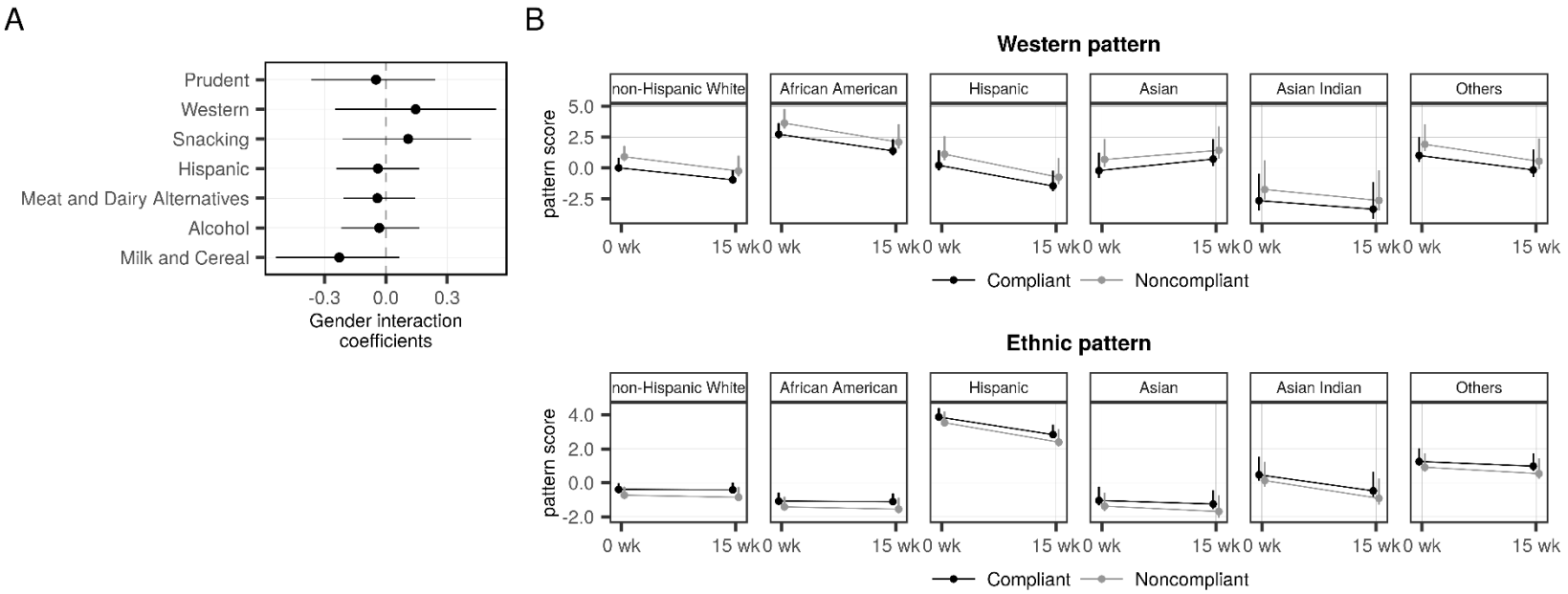


Figure 4.5: Influence of gender and race/ethnicity on exercise treatment effect on dietary preferences. Gender and race/ethnicity were dummy-coded where the reference groups were male and non-Hispanic white, respectively. Estimates were summarized using their posterior means with 95% credible interval. (A) Influence of gender on the treatment effect of exercise training. (B) Changes of the Western and Ethnic pattern scores by race/ethnicity.

4.4 DISCUSSION

In this study, we examined the influence of exercise training on habitual dietary preferences of young adults and its associations with exercise characteristics, including exercise duration, intensity, and dose. To describe overall dietary preferences, we used data-driven dietary patterns identified using a Bayesian sparse latent factor modeling. Seven dietary patterns were identified, including Prudent, Western, Snacking, Ethnic, Meat and Dairy Alternatives, Alcohol, and Milk and Cereal. Each dietary pattern score reflects an individual's relative standing regarding their frequency of making choices on a given dietary pattern. Changes in dietary preferences were evaluated before and after the 15-week exercise training protocol.

The current literature primarily focuses on the role of exercise in contributing to a negative energy balance via increasing energy expenditure, whereas there is still some controversy regarding the role of exercise on food intake (Blundell et al., 2015; Donnelly et al., 2014). At a physiological level, it has been established that exercise can influence appetite-regulating hormones (Stensel, 2010). However, such hormonal changes in response to exercise are not necessarily linked with post-exercise food intake because exercise may influence or be influenced by other behavioral, and environmental factors that are thought to affect their dietary intake. In the present study, most of dietary pattern scores were decreased after exercise training, which was consistent with the overall

decreasing trend in intake frequency of a wide array of food items. Although the 15-week exercise training did not result in substantial mean changes in anthropometric measures in the total cohort, the more positive changes in body/size were observed in those were exercise compliant, indicating that engaging in regular exercise may influence participants to regulate food intake as a means of controlling their body shape or improving health.

The influence of exercise training on dietary preferences was dependent on different components of exercise. We observed that a longer duration of exercise provides a motive for restraining dietary preferences in the Snacking and Western patterns, while a higher intensity of exercise induces an increased preference for the Prudent pattern. Although the current analyses linked exercise duration and intensity to different aspects of dietary preferences, the overall trend indicated an increased preference for healthier diets. This tendency was more pronounced in exercise dose where individuals with a higher exercise dose were more likely to prefer the Prudent pattern, while less likely to favor the Snacking pattern. These results support previous findings that exercise motivates individuals to pursue healthier nutritional habits (Blakely, Dunnagan, Haynes, Moore, & Pelican, 2004; Nigg et al., 1999; M. Tucker & Reicks, 2002). It is important to acknowledge that that there might be people who are vulnerable to failure in self-regulation of food intake. Those individuals may exhibit increased consumption of palatable but unhealthy foods following exercise, on the basis of

compensatory health beliefs (Knäuper, Rabiau, Cohen, & Patriciu, 2004). Thus, future research should address not only linear relationships between exercise characteristics and dietary responses, but also investigate the optimal level of exercise training that can lead to positive dietary changes, accounting for individual variability in response to exercise. Exercise duration was also linked with the increased preference for the Milk and Cereal pattern, but it was not possible to conclude its clinical implication due to the limited information on types and compositions of cereals that participants consumed.

When transitions of dietary preferences were compared between non-compliant and compliant individuals, those who were non-compliant showed a steeper decrease in the preference for the Prudent pattern and an increased preference for the Snacking pattern. Those who were compliant with the exercise prescription had a more moderate decrease in the preference for the Prudent pattern and a reduced preference for the Snacking pattern following exercise training. For young adults, snack foods are cheap, easy to access (e.g., off-campus grocer, on-campus cafeteria, and vending machines), and easy to store, compared to healthier options, such as fruits and vegetables. Indeed, students may not have the necessary skill set for planning their meals, and the college years have been associated with skipping meals, diminished food variety, and frequent consumption of salty and sugary snacks (Plotnikoff et al., 2015; Schweitzer, Ross, Klein, Lei, & Mackey, 2016; Strong, Parks, Anderson, Winett, & Davy, 2008).

The reduced preference for the Snacking pattern following exercise is parallel with the findings from previous research claiming that exercise may reduce snacking urges (Oh & Taylor, 2012, 2014; Taylor & Oliver, 2009; Thayer et al., 1993). Given that a majority of snack foods are high in sugar, sodium, and/or fat and snacking often occurs in the absence of hunger due to external factors such as pleasure, taste, and convenience (Hess, Jonnalagadda, & Slavin, 2016; Marquis, 2005; Roininen & Tuorila, 1999), a reduced urge to eat snack foods following exercise could have a significant impact on health.

Changes in dietary preferences in response to exercise varied by race/ethnicity. Hispanic participants showed a greater decrease in the preference for the Ethnic pattern. As indicated in the Ethnic pattern score at the baseline, this may simply due to their higher habitual consumption of foods belonging to the Ethnic pattern that consist of a group of Hispanic foods. It would have been easier for Hispanic participants to decrease consumption of food that they consumed more frequently. Asian participants exhibited a slight increase in the preference for the Western pattern following exercise training. It is not possible to explicitly identify the underlying mechanism and clinical significance of these observations, justifying more research on this topic. We did not identify a role of gender in influencing post-exercise dietary preferences, unlike previous animal studies that suggest sex-dependent effects on dietary preferences (Lee et al., 2017; T. Yang et al., 2017).

The present study has some limitations. First, the current analyses relied on the participants' self-report dietary intake, that may be subject to measurement errors. Second, there might have been a selection bias since students voluntarily participated in the study for which they received college course credit, within a limited course load. The participants of this study might have been more interested in improving their health behaviors. Finally, additional follow-up measurements were not collected after the end of study, making it difficult to evaluate whether or not the observed transitions persist. It may be possible that a participant engages in compensatory eating behaviors after the completion of intervention program, nullifying the observed positive trend for health brought on by exercise training.

In conclusion, the results of present study revealed that the15-week exercise training program motivated young adult participants to change their dietary preferences. The exercise-induced changes were associated with regulating food intake as well as encouraging healthier dietary preferences. Such a sequential relationship may assist individuals who feel pressured to improve their diet and physical activity at once, since imposing multiple goals often discourages, rather than motivates people to improve their health behaviors within a busy schedule. More research is required to better elucidate clinical significance and underlying mechanisms involved in the exercise-induced changes in dietary preferences.

Chapter 5. Genetics of Dietary Patterns among College Students

ABSTRACT

Background/Objectives: Among the many factors that influence eating behavior, genetic variation has consistently been shown to play a role, but few studies have examined genetic variation on a genome-wide scale. This study aimed to examine the contribution of genetic variation to dietary patterns among college students to better understand the determinants of eating behavior.

Methods: This study utilized data collected from 1,591 non-Hispanic white, African American, and Hispanic college students, who were defined as unrelated individuals based on their genomic relationships. Genome-wide single nucleotide polymorphism (SNP) data were used to estimate the contribution of genetic variation to the phenotypic variance of seven dietary patterns obtained from dietary pattern analysis based on self-administered food frequency questionnaire (FFQ) data.

Results: The estimated SNP-based heritability of dietary patterns ranged from 0.09 for the Alcohol pattern to 0.46 for the Snacking pattern. The SNP-based analyses explained a substantial portion of the total variance of the Western (0.42; 95% credible interval: 0.24, 0.60) and Snacking (0.46; 95% credible interval: 0.28, 0.64) patterns. In contrast, the Prudent pattern, which contained primarily

fruits, vegetable, and low-fat dairy foods, showed a lower heritability (0.14; 95% credible interval: 0.05, 0.24).

Conclusion: Genetic influences on dietary patterns among college students were apparent, but the range of the heritabilities across dietary pattern suggests that environmental exposures may also contribute substantially, especially for healthy foods. Increasing the accessibility and availability of healthy foods may be important in crafting dietary recommendations and interventions in this population.

5.1 INTRODUCTION

Understanding drivers of healthy and balanced eating is central to prevent and reduce the risk of obesity throughout all stages of the lifespan. In humans, substantial variability exists in dietary phenotypes, and research efforts have been made to dissect the sources of this variation and to examine why there are huge interindividual differences in response to dietary interventions and treatments (Rankinen & Bouchard, 2006). Twin and family studies have been used to establish heritability of dietary phenotypes (Pallister et al., 2014; Rankinen & Bouchard, 2006; Reed, Bachmanov, Beauchamp, Tordoff, & Price, 1997). In most studies, genetic effects accounted for variation in multiple diet-related phenotypes (e.g., energy intake, macronutrient selection, meal size, and meal frequency), albeit with varying heritability estimates across different age groups (Grimm & Steinle, 2011; Keller, Pietrobelli, Must, & Faith, 2002; Pallister et al., 2014; Rankinen & Bouchard, 2006). The substantial heritability of multiple dietary phenotypes suggests that robust prediction of genetic predisposition to dietary phenotypes will be of value in the development of effective clinical practice and dietary recommendations.

Previous studies has suggested that dietary patterns can be a useful means of representing phenotypes of overall dietary intake, accounting for interrelationships among multiple dietary components (Hu, 2002). Dietary patterns represent combinations of foods and beverages that people habitually

consume together and may be more predictive of overall health outcomes (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2015). Until now, a limited number of studies have examined the heritability of dietary patterns and have shown that a substantial proportion of phenotypic variance can be explained by genetic variation for both healthy and unhealthy types of dietary patterns, ranging from 0.12 to 0.79 in children and 0.33 to 0.54 in adults, despite distinct dietary patterns identified in each study (Bree et al., 1999; Breen et al., 2006; Faith et al., 2008; Gunderson et al., 2006; Keskitalo et al., 2008; Teucher et al., 2007; van den Berg et al., 2013). However, knowledge about the genetic contribution to dietary patterns among college students is still lacking in the literature. The college years are an important transitional period associated with various lifestyle changes (Gordon-Larsen et al., 2004). Given the unhealthy changes in eating behaviors and weight gain often observed in this life cycle period (Deliens et al., 2014), exploring the heritability of dietary patterns in college students may have significance for understanding genetic risk for unhealthy dietary behaviors and establishing appropriate environments for diet and weight management to prevent obesity, which may be less malleable in adulthood.

Recently, the literature suggests that the heritability of human traits can be estimated using genome-wide SNP data that enables the estimation of genetic associations for unrelated individuals (Campos, Vazquez, Fernando, Klimentidis,

& Sorensen, 2013; J. Yang et al., 2010). Indeed, previous studies have shown that a large proportion of the estimated heritability for quantitative traits (e.g., height) from twin studies can be captured using all the common SNPs concurrently (J. Yang et al., 2010; J. Yang, Manolio, et al., 2011). SNP-based heritability can provide a complementary approach for studying genetic predisposition to dietary traits, as it is not dependent on the assumptions of the family-based methods (e.g., equal environments assumption) that are often difficult to verify (Barnes et al., 2014). SNP-based heritability can also give some indication of which traits will be fruitful in future genome-wide association studies (GWAS). The estimated heritability derived from SNP-based genetic variation can indicate whether variation in a given population is likely to be explained by GWAS analyses (J. Yang et al., 2017). Although several candidate gene association studies revealed risk alleles in/near obesity susceptibility genes (e.g. *FTO*) that influence a multitude of dietary phenotypes (Cecil et al., 2008; J. M. McCaffery et al., 2017; Jeanne M. McCaffery et al., 2012; Speakman, Rance, & Johnstone, 2008; Timpson et al., 2008; Wardle et al., 2008), there might be a large number of common variants with small effects yet to be identified in other regions. However, this approach has not been applied to the evaluation of the genetic underpinnigs for human eating patterns.

In this study, we aimed to investigate SNP-based heritability of dietary patterns among college students using data from the Training Intervention and

Genetics of Exercise Response (TIGER) study. To estimate the genetic contribution to variation in dietary patterns, we used genome-wide SNPs obtained from unrelated individuals in the study population. We also examined the heritability estimates of consumption for food groups, including vegetables, fruits, meat, dairy products, grains, fats and oils.

5.2 METHODS

5.2.1 Study population

Data used in these analyses were drawn from the TIGER study, which aimed to investigate the genetic underpinnings of response to exercise training and introduce college students to regular exercise via a 15-week aerobic exercise training protocol. The target subjects were sedentary college-age students who did not engage in caloric restriction and exercised less than 30 min/wk during the 30 days prior to study enrollment. Fifteen multi-racial cohorts were recruited from the University of Houston (2003-2008) and the University of Alabama at Birmingham (2010-2015). The exclusion criteria for study participation entailed physical contraindication to exercise, pregnancy, and/or a known metabolic disorder that may impact body composition. The TIGER study was approved by each of the participating Institutional Review Boards, and informed consent was obtained from all study subjects. More details regarding the TIGER study design are available elsewhere (Miller et al., 2014; Sailors et al., 2010).

5.2.2 Dietary assessment and dietary patterns

Dietary intake at baseline was evaluated using a self-administered Block Food Frequency Questionnaire (FFQ, NutritionQuest, Berkeley, CA). Habitual consumption of 102 food items was obtained via nine frequency categories, ranging from "never" to "every day", which were then converted to daily frequencies. The FFQ data also included daily servings of six food groups, including vegetables, fruits, meat, dairy products, fats and oils, and grains. Bayesian sparse latent factor analysis was applied to the dietary data obtained from the subset of study participants who completed at least 90 % of food item questions and had valid observations for demographic and anthropometric information (n = 2,730). Dietary patterns were annotated based on their food subsets determined by posterior inclusion probabilities, a measure of how much the model favors the inclusion of a food in each dietary pattern. Seven dietary patterns were generated from the FFQ data in previous analyses, which were denoted Prudent, Western, Snacking, Ethnic, Meat and Dairy Alternatives, and Milk and Cereal. Factor scores that represent an individual's position on each dietary pattern were used as quantitative measures of each dietary pattern.

5.2.3 Genotyping

Genotyping was carried out using the Metabochip (Illumina, San Diego, CA), a custom genotyping array that interrogates genotypes for 196,725 SNPs relevant to metabolic, cardiovascular, and anthropometric traits (Voight et al.,

2012). The accuracy and quality for the genotyping was examined by assessing the consistency between technical replicates. The average percentage of SNP genotype mismatch between technical replicates was 0.31% ± 0.13. To assess the performance of genotyping calling, we randomly assigned DNA samples from 39 participants to different plates and examined the proportion of mismatch in genotyping calling. The average mismatch rate of SNP genotypes for the pairs was 0.17% ± 0.12 (mean ± SD).

5.2.4 Exclusion of samples and SNPs

For SNP-based heritability analysis, this study focused on a subset of study participants that included non-Hispanic white, African American, and Hispanic individuals. For subjects who were related as determined by genomic relationships estimated from genetic markers (> 12.5% genotype sharing), a single sample was chosen to represent the relative pair. In total, 1 591 subjects were available, which included 799 non-Hispanic whites, 514 African Americans, and 278 Hispanics.

SNPs were excluded based on the following criteria: technical quality control failure based on Illumina GenCall score, monomorphic SNPs, SNPs with a minor allele frequency (< 3%), and SNPs with replicate errors between technical replicates. As a consequence, 109,138 SNPs, 124,875 SNPs, and 109,562 SNPs were retained within non-Hispanic white, African American, and Hispanic datasets, respectively.

5.2.5 Statistical analysis

Average daily servings of food groups were summarized using mean and standard deviations (SD). The differences in daily servings between female and male were evaluated ANCOVA, adjusted for age, race, and BMI. The Genomic Best Linear Unbiased Predictor (G-BLUP) model, a commonly used whole genome regression method for human data, was applied to the present study in a Bayesian context in order to estimate SNP-based heritability (h^2_{SNP}) (Campos et al., 2013; J. Yang et al., 2010). First, genomic relationships that measure genetic similarity between individuals were inferred using genome-wide samples of SNPs in which a genomic relationship matrix was calculated using cross-products between marker genotypes. SNP-based heritability was then calculated as the ratio of variance of phenotypes explained by this relationship matrix, based on a linear model. This is equivalent to estimating the genetic variance by fitting all common SNPs used to measure genomic relationships, simultaneously (Vinkhuyzen, Wray, Yang, Goddard, & Visscher, 2013; J. Yang et al., 2010).

A cube root transformation was applied to the factor scores for the Alcohol and Meat and Dairy Alternatives patterns in order to lessen the high positive skewness. Factor scores for each dietary pattern were used to estimate SNP-based heritability of each dietary pattern controlling for covariates, including age, gender, and data collection site. The analyses also included the first six principal components obtained from the genotype data to account for the genetic

structure of the study population. All analyses were performed on the total sample first, and then on women and men separately, using R version 3.4.3 (R Core Team, 2017).

5.3 RESULTS

Table 5.1 displays the average daily servings of major food groups of study subjects. Participants' mean age was 21.52 y (SD = 3.24) and two thirds of subjects were females (64.24 %). Male subjects exhibited greater consumption of meat, dairy product, and grain groups and showed less consumption of fats and oils, compared to females.

Figure 5.1 shows the SNP-based heritability estimates of dietary patterns, which represent the proportion of phenotypic variance explained by common SNPs in the sample. The values were summarized with 95% Bayesian credible intervals that describe a range having a 95% probability of containing the true value of heritability. The heritability estimates ranged from 0.09 for the Alcohol pattern to 0.46 for the Snacking pattern. The Prudent pattern, which represents healthier types of foods, showed a lower heritability (h^2_{SNP}: 0.14), compared to the Snacking (h^2_{SNP}: 0.46) and Western (h^2_{SNP}: 0.42) patterns, which generally represent unhealthier types of foods, suggesting patterns for these foods may be strongly influenced by gene-based biological variability. The Ethnic pattern, which consists primarily of Hispanic foods, had a heritability of approximately

0.27. The heritability estimates of the Meat and Dairy Alternatives, and the Milk and Cereal patterns were 0.10 and 0.18, respectively. When SNP-based heritability was estimated by gender, some dietary patterns showed differences between females and males. For the Prudent pattern, the genetic contribution to the total variation was higher in the male group (h^2_{SNP}: 0.29) than the female group (h^2_{SNP}: 0.18). For the Meat and Dairy Alternatives pattern, the heritability estimate of the male group was 0.20, which was almost double compared to that of the female group. In contrast, the lower heritability for the Western pattern was estimated in the male group (h^2_{SNP}: 0.39) than the female group (h^2_{SNP}: 0.49).

Figure 5.2 displays the heritability estimates for consumption of FFQ food groups. The heritability estimates for those groups ranged from 0.11 for dairy products to 0.25 for fats and oils. This trend is similar to what was observed in dietary patterns. The proportion of genetic effect on total phenotypic variation in vegetables (h^2_{SNP}: 0.12) and fruits (h^2_{SNP}: 0.15) are comparable to the heritability estimate of the Prudent pattern (h^2: 0.14). The heritability estimates of meat (h^2_{SNP}: 0.19), grains (h^2_{SNP}: 0.21), fats and oils (h^2_{SNP}: 0.25), which were major elements of the Western and Snacking patterns, were higher than those of vegetables and fruits, similar to the higher heritability of the Western and Snacking patterns, compared to the Prudent pattern. Additionally, the male group showed a higher heritability for vegetables, fruits, and dairy products, while showing the lower heritability for grains, compared to the female group.

Table 5.1: Characteristics of the study subjects.

	Female (n = 1,022)	Male (n = 569)	Total (n = 1,591)
Age (year)	21.52 (3.24)	21.39 (3.24)	21.74 (3.22)
Racial/ethnic groups (n)			
Non-Hispanic white	477	322	799
African American	388	126	514
Hispanic	157	121	278
Daily servings of food group		Mean (SD)	
Vegetables	2.15 (1.82)	2.09 (1.64)	2.13 (1.76)
Fruits	1.21 (0.91)	1.17 (0.93)	1.19 (0.92)
Meat*	1.93 (1.44)	2.95 (1.92)	2.30 (1.70)
Dairy products*	1.04 (0.84)	1.45 (1.22)	1.19 (1.01)
Grains*	4.40 (2.84)	5.51 (3.30)	4.80 (3.06)
Fats and oils*	3.13 (1.84)	2.90 (1.69)	3.05 (1.80)

* Males significantly different from females (P-value < 0.05 from one-way ANOVA adjusting for age, race and BMI)

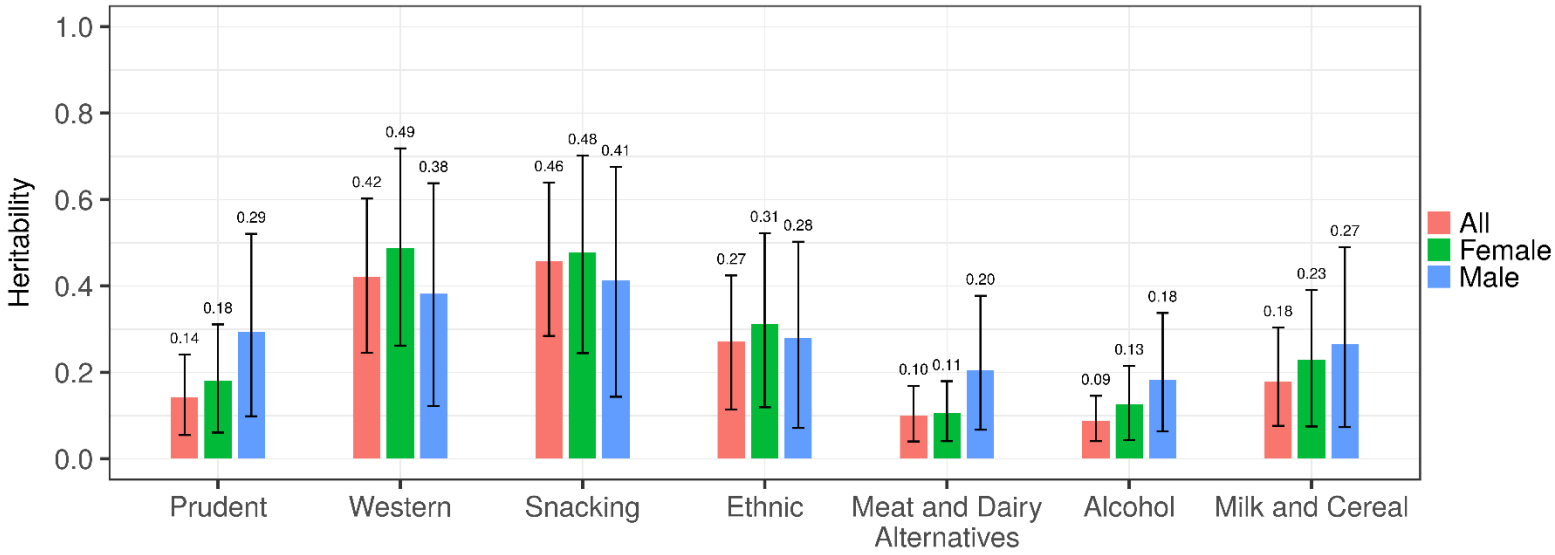


Figure 5.1: Heritability estimates for habitual dietary patterns among college students with 95% Bayesian credible intervals. The SNP-based heritability was examined on total subjects and by gender using the genotype data from unrelated individuals.

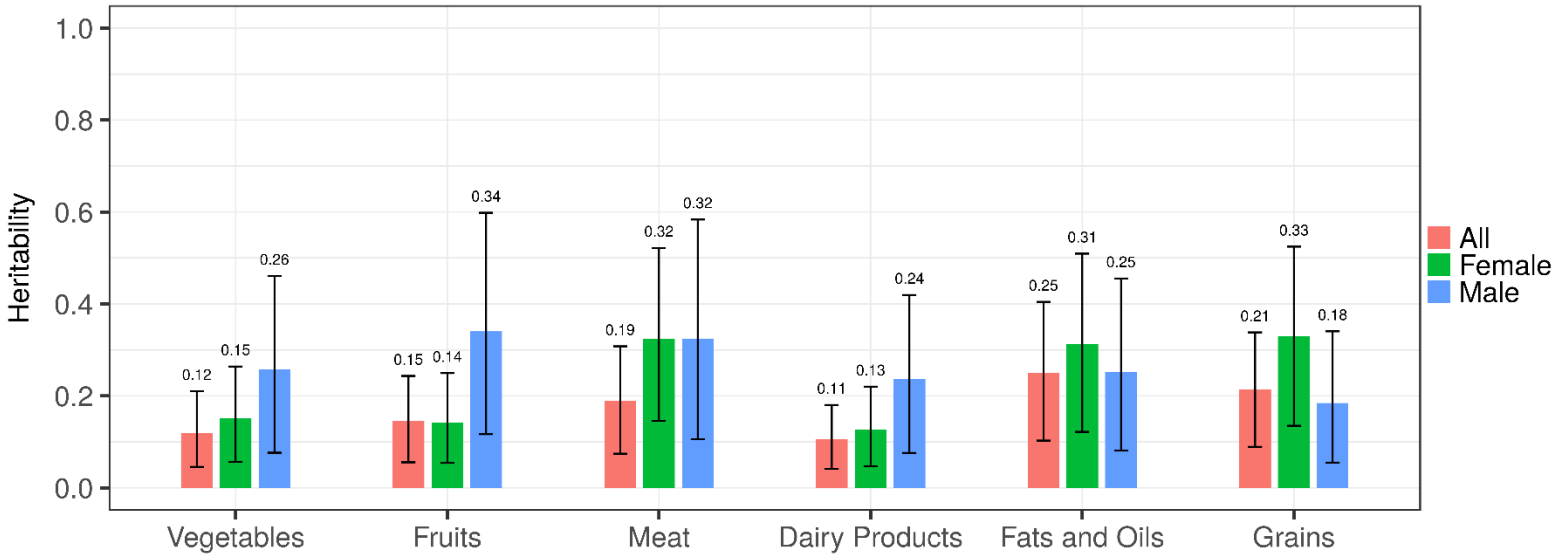


Figure 5.2: Heritability estimates for food group intakes among college students with 95% Bayesian credible intervals. The SNP-based heritability was examined on total subjects and by gender using the genotype data from unrelated individuals.

5.4 DISCUSSION

In the present study, we examined the heritability estimates of dietary patterns identified among college students using the data from the TIGER study. We used the genotyped SNPs from unrelated individuals to evaluate the SNP-based heritability of dietary patterns, which provided heritability estimates based on genetic data, as opposed to estimates based on relatedness as in family study designs (Bree et al., 1999; Breen et al., 2006; Faith et al., 2008; Gunderson et al., 2006; Keskitalo et al., 2008; Teucher et al., 2007). This alternative approach is useful for studying genetic predisposition to dietary behaviors in unrelated cohorts.

Our results showed that genetic variation accounts for more than 40% of the total variance in the Snacking and Western patterns, which mainly included foods high in sugar, salt, and fat. This finding confirms the existing literature that displayed a sizable genetic influence on unhealthy types of dietary patterns (Bree et al., 1999; Keskitalo et al., 2008). In contrast, genetic variation explained only a small portion of total variance in the Prudent pattern, which was generally represented by frequent consumption of vegetables, fruits, and low-fat dairy foods. The higher heritability of the Snacking and Western patterns may be explained in part by the fact that humans have been evolved to seek and eat energy dense foods, leading to greater taste preferences for palatable foods that are high in fat and sugar (Breslin, 2013). For example, It has been suggested that

genetic variation in the CD36 fat taste receptor, T1R2 sweet taste receptor, and T2R38 bitter taste receptor may confer genetic predisposition to the excessive intake of unhealthy foods (Chamoun et al., 2018). In addition to the genetic effects of taste, individuals with obesity risk variants at fat mass and obesity-associated gene (*FTO*) have greater caloric and fat intake, and exhibited preference for high-fat foods (Cecil et al., 2008; Speakman et al., 2008; Timpson et al., 2008; Wardle et al., 2008). *FTO* variants were also related to more frequent consumption of snacks among overweight or obese individuals.(Jeanne M. McCaffery et al., 2012) In the EPIC study, common allelic variations in leptin gene (*LEP*) were associated with extreme snacking behavior in women (Krom et al., 2007). In the Nurses' Health Study, a common obesity variant near melanocortin 4 receptor gene (*MC4R*) was associated with greater energy intake in women (Qi et al., 2008). Obesity susceptibility loci in brain-derived neurotrophic factor gene (*BDNF*) and *MC4R* gene also related to greater energy intake among individuals at high risk of diabetes (J. M. McCaffery et al., 2017).

Genetic variation has also been reported to predispose individuals to avoid the bitter taste of healthy foods, including specific vegetables and fruits (e.g., cruciferous vegetables) (Breslin, 2013; Chamoun et al., 2018; Grimm & Steinle, 2011), although we observed a relatively low heritability for the Prudent pattern. While the low heritability does not mean that genetic variation is irrelevant, it may indicate that the environment related to foods associated with the Prudent

pattern was a more significant contributor to consumption. Compared to energy-dense foods, healthier types of foods, such as vegetables and fruits, are more expensive, which may discourage college students living on their own from eating such foods when they are financially limited, even if they have a preference for eating such foods (Drewnowski & Darmon, 2005). Vegetables and fruits have a limited accessibility among college students, compared to snacks and fast foods, which can be easily obtained from on-campus cafeteria or vending machines (Strong et al., 2008). The accessibility of healthy foods can also be different by residence type, as demonstrated in previous research in which students residing in residence halls eat more fruits and vegetables and consume more various dairy products than those living off-campus (Brunt & Rhee, 2008). Taken together, lifestyle transitions in college students might have had a more profound effect on the inequality of environmental exposures involving healthy eating patterns, resulting in a lower contribution of genetic influence for the Prudent pattern than for the Snacking and Western patterns.

The heritability of the Alcohol pattern was lower, compared to other dietary patterns. Alcohol consumption is known to be determined to some extent by alcohol metabolizing genes (Clarke et al., 2017; Gelernter et al., 2014). To date, studies on genetic predisposition to alcohol behavior have been focused on alcohol use disorder whose heritability estimate was approximately 0.5 in the recent meta-analysis of twin and adoption studies (Verhulst, Neale, & Kendler,

2015). Evidence for heritability of habitual alcohol consumption is relatively scarce, but Clarke et al. (Clarke et al., 2017) recently provided a SNP-based heritability estimate of self-reported alcohol consumption using 89,175 white British individuals in the UK Biobank sample. In that study, the estimated heritability of alcohol consumption in total sample was 0.13, in males was 0.15, and in females was 0.13. This is similar to our finding in which the heritability for the Alcohol pattern was estimated to be 0.09 in total sample, 0.18 in males, and 0.13 in females, although the UK Biobank sample consisted of individuals aged 40-69 years (Sudlow et al., 2015).

When SNP-based heritability was examined within each gender, the relative genetic contribution to the phenotypic variation was heterogeneous in some dietary patterns. Males displayed a higher heritability estimate for the Prudent and Meat and Dairy Alternatives pattern, while showing a lower heritability for the Western pattern, compared to females. This could imply that genetic variation may exert different influences on eating patterns. Dietary behavior is a complex trait influenced by many different factors, and there are multiple paths to reach specific dietary phenotypes. Gender differences in heritability for dietary patterns could have arisen if the relative prevalence of these paths was dependent on gender. Additionally, gene-gene and gene-environment interactions could be a potential source of gender difference in the heritability estimates. It is possible that expression of genes involved in specific

dietary traits is modulated by gender-specific features in physiological functions controlled by other genetic components. Activation of casual genes and the related pathways for dietary behaviors can also be influenced by environmental factors, such as exercise habits, for which female and male differ (Craft, Carroll, & Lustyk, 2014). However, replication is required with the same study population to confirm the observed trend, given the uncertainty intervals of gender-specific heritability estimates.

The heritability estimates of food groups displayed a similar trend seen in dietary patterns. The estimated heritability of vegetables and fruits, which are a major component of the Prudent pattern, were lower than those of meat, grains, and fats and oils, which are mainly associated with the Western and Snacking pattern. The estimated heritability for vegetables and fruits were higher in males than females, which is analogous to what was seen in the Prudent pattern.

This study has some limitations. First, the dietary data were obtained based on self-reported food frequencies that may poorly represent dietary phenotypes due to measurement errors and potential underreporting among certain subjects, particularly those with high alcohol intake (Lutomski, van den Broeck, Harrington, Shiely, & Perry, 2011). Second, our study samples were restricted to non-Hispanic whites, African Americans, and Hispanics. Thus, study findings are limited to those populations. A study investigating other race groups

may yield different results. Finally, the heritability values could have been underestimated due to the limited number of SNPs genotyped on the Metabochip.

In conclusion, we examined SNP-based heritability of habitual dietary patterns among college students. The proportion of phenotypic variance attributable to genetic variation varied by dietary patterns and food groups. We observed that healthier types of dietary patterns and food groups have a relatively lower heritability than unhealthy ones. Such differences would suggest that environments involving healthy diets are less uniform among college students. Increasing the accessibility and availability of healthy foods may have important implications for designing intervention strategies and public policy to improve overall health. Importantly, the low values of heritability estimates observed in some dietary patterns and food groups are not a demonstration that genetic variation plays little role in the phenotypes of dietary intake. Future studies of the underlying genetic pathways that drive the phenotypic difference in eating patterns would greatly inform future dietary interventions.

Chapter 6. Conclusion

This dissertation aimed to investigate the effects of genetic variation and exercise on patterns of habitual food intake in young adults. To this end, statistical modeling of dietary patterns from the Bayesian perspective was used to create a meaningful overall representation via latent patterns of food intake in observed data (Chapter 3). This research has established that individuals who are engaging in regular exercise tended to alter their diet towards healthier eating patterns with lower consumption of snacks and unhealthy foods (Chapter 4). Finally, the SNP-based heritabilities of dietary patterns and food group intakes were estimated via whole-genome regression based on the gnome-wide SNP data (Chapter 5).

Dietary pattern analysis. One major contribution of this research is the application of a flexible, probabilistic model for characterizing dietary patterns. Since foods are consumed in a variety of combinations in free-living conditions, analyses on a per-food basis of disease risk and/or associations with phenotypic traits may be misleading. Dietary pattern analysis offers a reasonable alternative to circumvent this challenge by capturing overall food consumption behaviors with a small number of latent features, accounting for inherent interrelationships among multiple foods and nutrients. Common methods of dietary pattern analysis such as PCA and EFA have several limitations, chief among which are the arbitrary *post hoc* decisions made by investigators in several intermediate stages.

Traditionally, *post hoc* decisions of dietary pattern analysis were involved in the factor rotation and loading truncation, which attempts to find a sparse factor orientation to simplify interpretation of dietary patterns. For this reason, Bayesian sparse latent factor modeling was utilized to overcome the problems of classical identification criteria via regularization that exploits sparsity-inducing priors on factor loadings as an integral part of the estimation of dietary patterns. The results from our empirical study demonstrated that sparsity-inducing priors can provide a more coherent procedure of deriving dietary patterns by naturally segregating the important food-pattern associations from those that have negligible associations based on statistical inference rather than on an investigator's intuition.

One of the principal advantages of the proposed alternative for dietary pattern analysis is its modularity and extensibility, allowing the dietary pattern model to be readily embedded into a more complex model. To accommodate covariates in the PCA/EFA-based dietary pattern modeling, we have relied on two-step regression adjustments, in which individual food intakes are regressed on a set of covariates to compute food intake residuals and these residuals are used as new input variables for dietary pattern analysis. Bayesian sparse latent factor analysis can be extended to overcome this limitation by coupling ANOVA and regression components based on covariates. Moreover, the proposed Bayesian approach has the ability to automatically respond to any shape of food intake distributions by introducing the Dirichlet process mixture model to the modeling

of factor score distributions. This is an improvement from the classical approach that primarily relies on the use of log transformation after removing zeros by adding an arbitrary constant, in order to approximate normal distributions.

Exercise and dietary patterns. Dietary patterns were derived using data from sedentary young adults who have undergone 15 weeks of aerobic exercise training, which were annotated as Prudent, Western, Snacking, Ethnic, Meat and Dairy Alternatives, Alcohol, and Milk and Cereal. An important goal of this project was to determine if and how dietary intake changes following the exercise intervention. The results have demonstrated that exercise can serve as a gateway behavior for the consumption of healthy foods among college students. A gateway relationship of behaviors represents one behavior that, if adopted by individuals, would change another behavior. This study hypothesized that participation in regular exercise will lead to dietary changes toward a healthier direction based on the positive relationship between physical activity and healthy food consumption such as vegetable and fruits, reported in the previous cross-sectional studies.

Dietary patterns were used to evaluate the changes in overall food consumption after 15-weeks of aerobic exercise training. One challenge was to measure changes in dietary patterns, as they are not directly observable and therefore, comparisons over time are not straightforward. Factor loadings are assumed to remain the same over time, which allows for the comparison of factors

scores at different times on the same basis. Accordingly, changes in dietary patterns were captured through the differences of factor scores estimated from dietary data collected before and after the exercise training. Additionally, these changes were examined in relation to different components of exercise, such as duration, intensity, and dose, because they may have a different degree of influence on dietary patterns.

Factor scores of most dietary patterns were decreased after the 15-week exercise training regimen. Given the decrease in several anthropometric measures, including weight and waist circumference in exercise compliant subjects, this result may suggest that individuals adopt subsequent regulation of food intake after initiating exercise. Individuals with a higher exercise intensity were more likely to favor the Prudent pattern, and those with a longer duration of exercise were less likely to prefer the Western and Snacking patterns. Consistent with this finding, individuals with a higher total dose of exercise exhibited a greater preference for the Prudent pattern while showing a lower preference for the Snacking pattern. Overall, the observed trends in dietary patterns suggested that engaging in regular exercise may encourage individuals to adopt a healthier diet.

Heritability of dietary patterns. The examination of the SNP-based heritability of food intake phenotypes was motivated by a hypothesis that these phenotypes are affected by numerous small-effect genes. To date, the functional contributions

of individual variants associated with food consumption to the total phenotypic variance have not been clearly elucidated and explain only a small proportion of the total variance, suggesting that much larger numbers of variants with smaller effects have not yet been discovered. Thus, it is important to establish the joint contribution of genetic variants to the variability of a food intake phenotypes to enhance our understanding of genetic underpinnings of a specific food intake phenotype.

The genetic relatedness of individuals using genome-wide SNPs was used to eliminate unknown replicates and reduce relative pairs to a single sample of representative individuals. Genome-wide genotypes in the remaining individuals were then used to examine the extent to which the phenotypic similarity across individuals could be accounted for by their genetic similarity, arriving at an estimate of SNP-based heritability for each dietary pattern and food group intake. This approach also allows SNPs with smaller effects, which may remain difficult to pinpoint in genetic association studies, to jointly contribute to explaining the variability of food intake phenotypes. Although the SNP-based heritability does not directly identify predictive variants, it tells us how well we could predict a trait from genetics by gauging the relative importance of genetic and environmental factors in a given population. Understanding the genetic underpinnings of a given trait provides some guidance for planning appropriate regulatory environments for a particular aspect of food consumption behavior and

related health outcomes. Importantly, SNP-based heritability can be estimated from a sample of unrelated individuals, and therefore, it is less likely to be biased by common shared environment effects.

The heritability estimate for the Prudent pattern was lower, compared to the Western and Snacking patterns in the sample of college students from the TIGER study. The heritability estimates for vegetables and fruits were small, compared to fats and oils. These results may indicate that, while biological factors may play a stronger role in driving preferences of sweet, salty, and high fat foods, environment may more strongly determine healthy eating behaviors. Many students are separated from their normal home life when they go to university, and acquire more autonomy over their dietary choices, but often are financially more limited than then living at home. College living for many students may lead to avoidance of healthy eating choices in an attempt to balance a food budget, since healthy foods are generally more expensive than unhealthy foods. Moreover, compared to healthy foods, unhealthy foods such as snacks and fast foods are more readily available at campus cafeterias or vending machines, and easy to store. Thus, environmental changes regarding transition to university might have had a stronger influence on the variability of healthy food consumption. Policies to help offset cost differences between healthy and unhealthy foods, and increased accessibility and availability of healthy foods for college students would be warranted, based on these findings.

Estimated heritability can help gauge which dietary traits would be more fruitful in future GWAS. It is unlikely to find predictive variants accounting for a sizable portion of the variability of food intake patterns in a given population if the estimated heritability is low. The estimated heritability may also help appraise population-level efficiency of genetic testing for predicting genetic predisposition toward a particular phenotype of food consumption and risk for nutrition-related diseases in a given population.

FUTURE DIRECTIONS

Potential future work directions from this dissertation project are outlined below.

Dietary pattern modeling. The applied approach in this dissertation still assumes the number of dietary patterns is known in advance. This assumption could be relaxed to allow the data to determine this number. The model can be further extended to include clinical response variables or disease states in which latent factors become regressors to predict health outcomes. The current methods, PCA and EFA, relies on a two-stage model for this task, in which factor scores (i.e., dietary pattern scores) are calculated from rotated loadings, and then plugged into a regression model for the examination of associations between dietary patterns and a health outcome of interest. Additionally, a general framework is needed to

model changes in dietary patterns from data measured at multiple times points, which are likely to have non-linear trajectories.

Exercise and dietary intake. The observed changes in dietary intake associated with exercise need to be examined for their stability over a longer duration. Compensatory changes in dietary intake may take place after a longer period, in which an individual is motivated to consume palatable foods based on belief that the negative effects of unhealthy eating can neutralized by the positive effects of exercise. Additionally, healthy dietary behaviors could be explored as a means to promote physical activity.

Genetics of food intake phenotypes. Future research will no doubt identify common, casual loci associated with food intake phenotypes. A phenotype that showed a higher heritability, such as snack consumption, would be a good starting point for future genetic studies. How exercise may alter a genetic predisposition to unhealthy dietary habits would comprise a fruitful area of intervention research. Finally, a better means of assessing food intake phenotypes is needed instead of relying on self-reported food consumption. Omics technologies may be a key to providing more reliable and accurate depiction of phenotypes.

Appendices

APPENDIX A

Sparsity-inducing priors for Bayesian analysis of sparse latent factor model used for deriving dietary patterns among the TIGER study participants.

Sparsity-inducing priors take the form (Carvalho et al., 2008; Lucas et al., 2006; West, 2003)

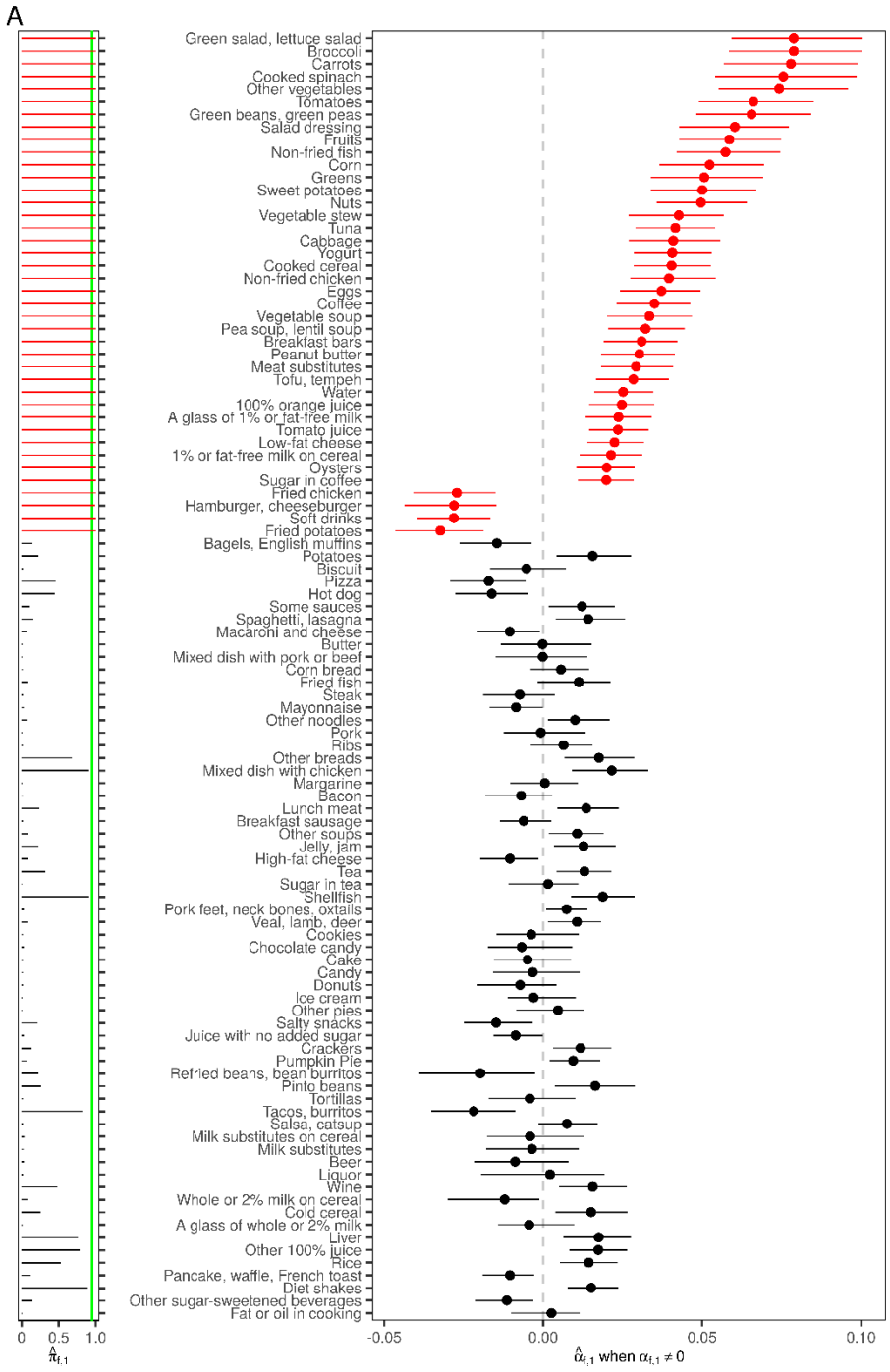
$$\alpha_{f,j} \sim (1-\pi_{f,j})\delta_0(\alpha_{f,j}) + \pi_{i,j}N(\alpha_{f,j} \mid 0, \tau_j)$$
$$\pi_{f,j} \sim (1-\rho_j)\delta_0(\pi_{f,j}) + \rho_j Be(\pi_{f,j} \mid a_j m_j, a_j(1-m_j))$$

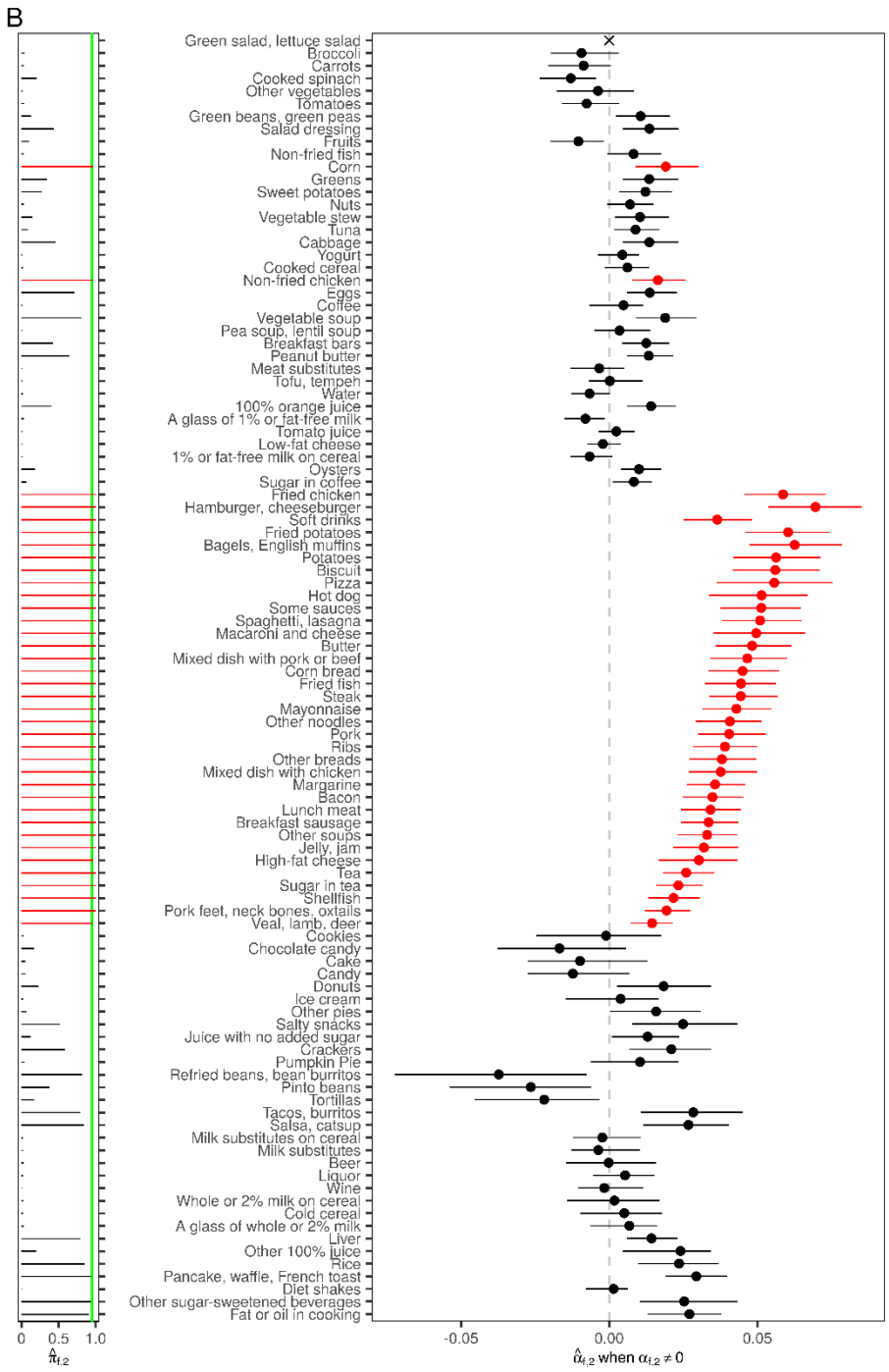
where $\alpha_{f,j}$ is the loading for food f associated with j-th dietary pattern; $\pi_{f,j}$ is the probability of that loading being non-zero; $\delta_0(\cdot)$ is a point-mass at zero; $Be(\pi_{f,j} \mid a_j m_j, a_j(1-m_j))$ is a beta distribution with pattern-specific mean m_j and precision parameter $a_j > 0$;. ρ_j is a pattern-specific probability assumed to be drawn from $\rho_j \sim Be(\rho_j \mid sr, s(1-r))$ that favors small values to encourage sparsity; ρ_j, a_j, and m_j control the distribution of j-th dietary pattern over the $\pi_{f,j}$. This prior allows exact zeros in the factor loading, allowing us to explain some of food-pattern associations by assuming $\alpha_{f,j} = 0$. This approach assumes that only a subset of foods plays a role in any dietary pattern, and thus each pattern defines a distinct cluster of interrelated foods. Each loading may be zero

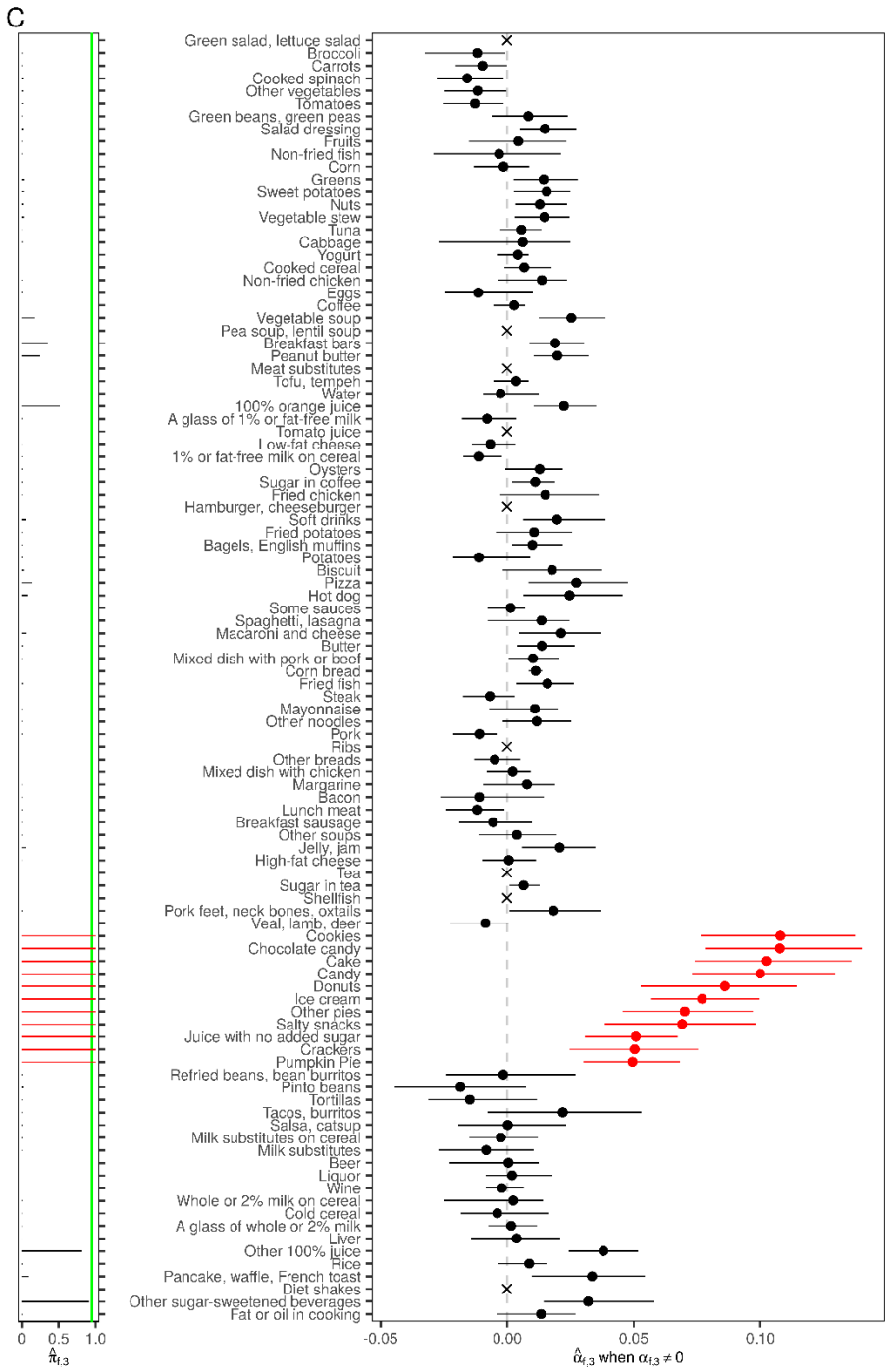
or have non-zero values drawn from Gaussian distributions, implying that the prior distributions of each loading have a mixture distribution made up of a point mass at zero and a Gaussian distribution over non-zero values as in standard Bayesian variables selection analysis (Carvalho et al., 2008; George & Clyde, 2004; Lucas et al., 2006).

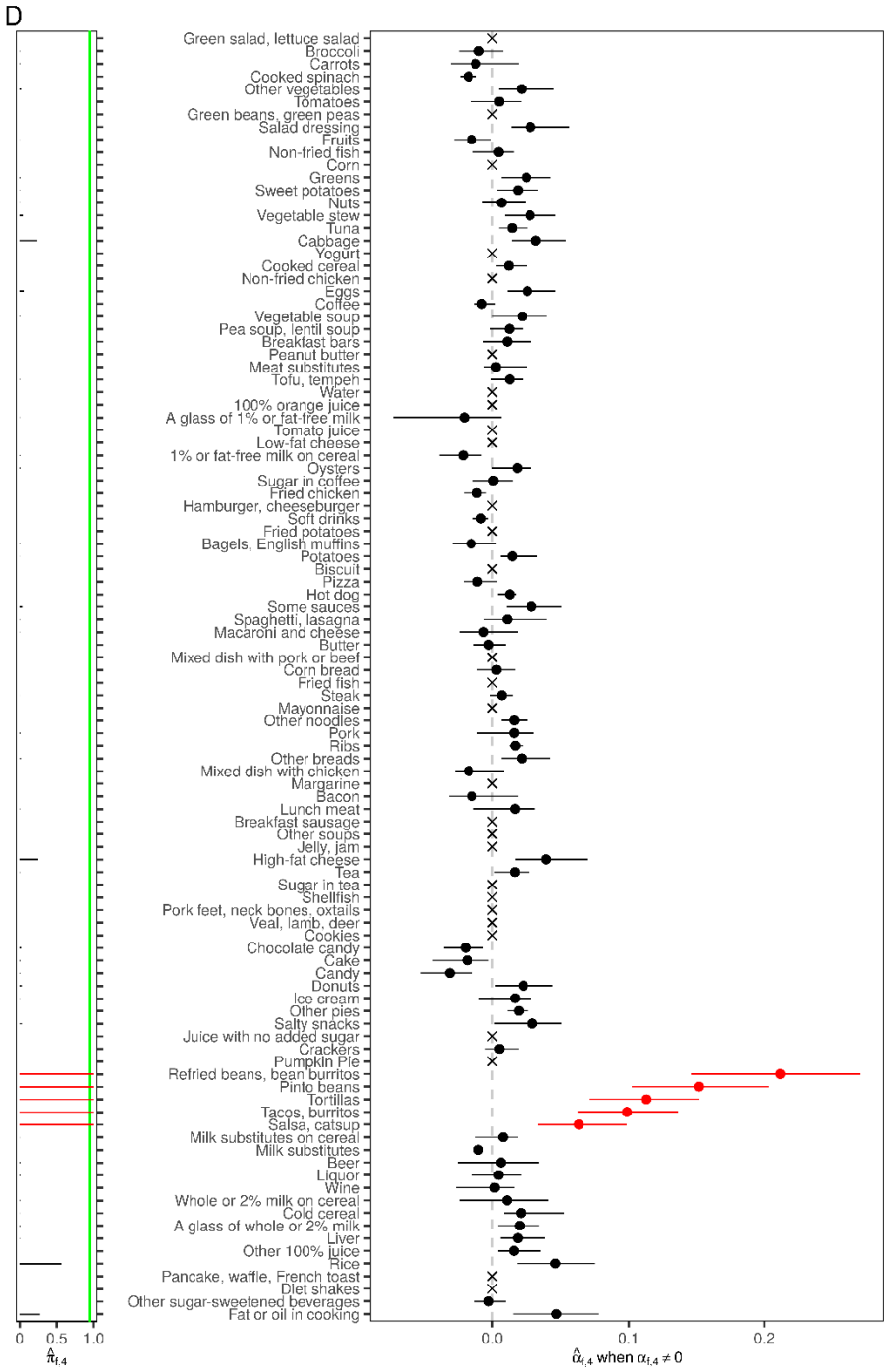
Based on this prior specification, and our data $x_{1:n}$, we can calculate the posterior distribution over the inclusion probabilities $\hat{\pi}_{f,j} = \Pr(\alpha_{f,j} \neq 0 \mid x_{1:n})$ that are central to defining a set of foods for specific dietary patterns. Importantly, the j-th pattern has its own common base-rate of non-zero loadings set at $\mathrm{E}(\pi_{f,j} \mid \rho_j) = \rho_j m_j$. This makes $\hat{\pi}_{f,j}$ shrink towards zero for foods having little association with the j-th dietary patterns while shrinking towards the estimated base-rate for foods showing a non-zero association. Such natural shrinkage facilitates the identification of significant food-pattern associations from the remainder of the food items.

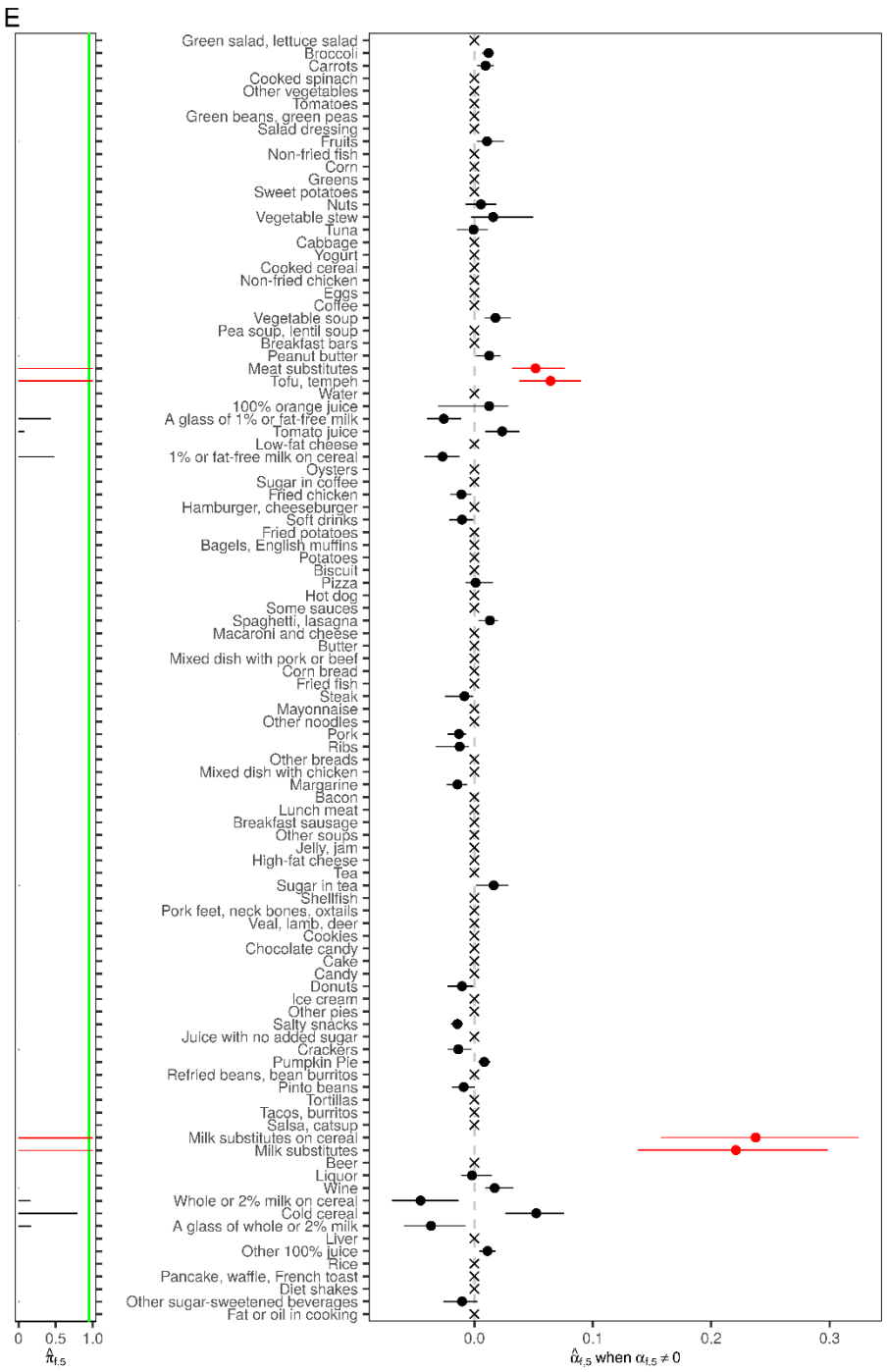
APPENDIX B


A
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,1}$
-0.05
0.00
0.05
0.10
$\hat{\alpha}_{f,1}$ when $\alpha_{f,1} \neq 0$

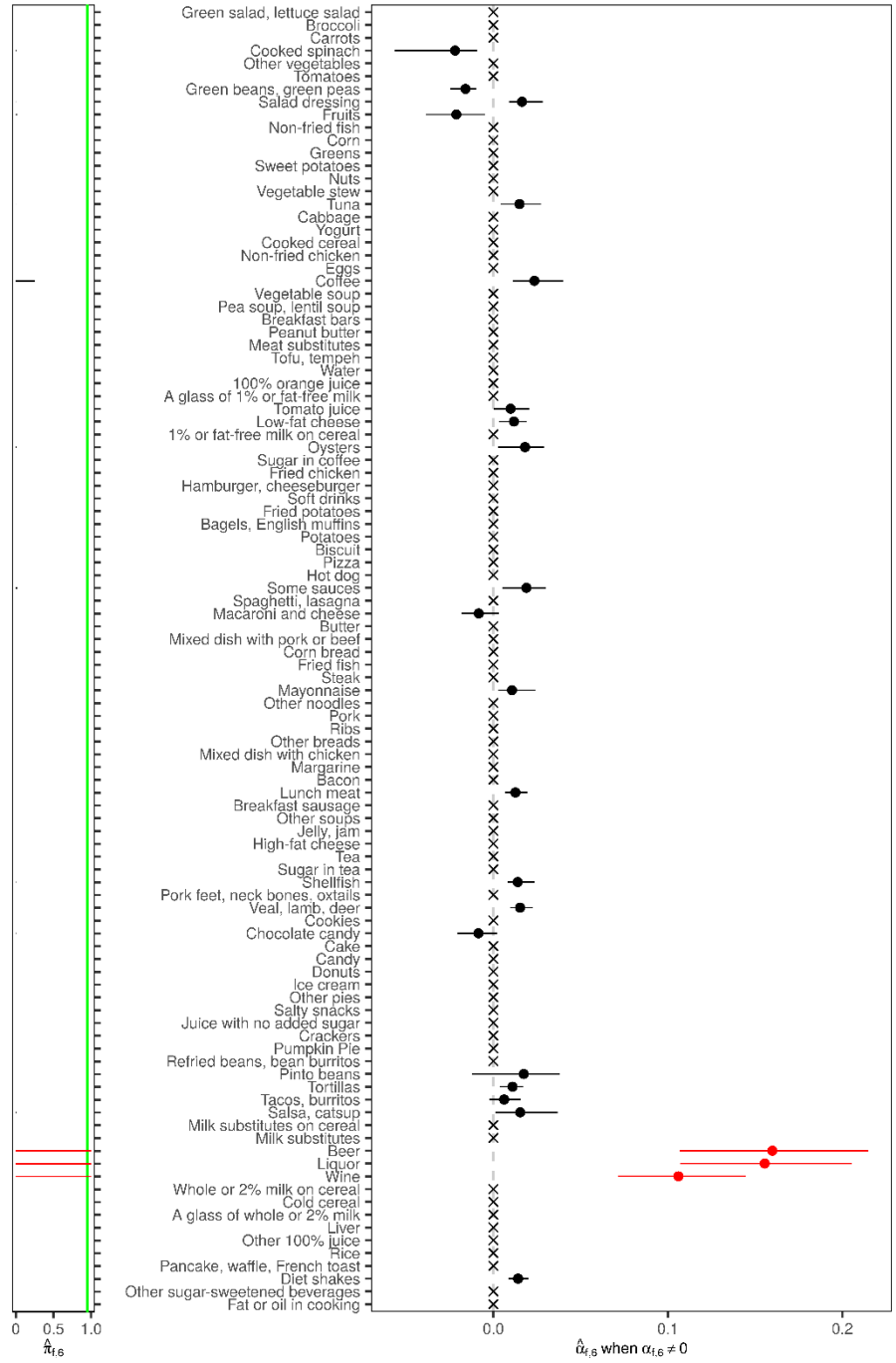

B
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,2}$
-0.05
0.00
0.05
$\hat{\alpha}_{f,2}$ when $\alpha_{f,2} \neq 0$


C
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,3}$
-0.05
0.00
0.05
0.10
$\hat{\alpha}_{f,3}$ when $\alpha_{f,3} \neq 0$

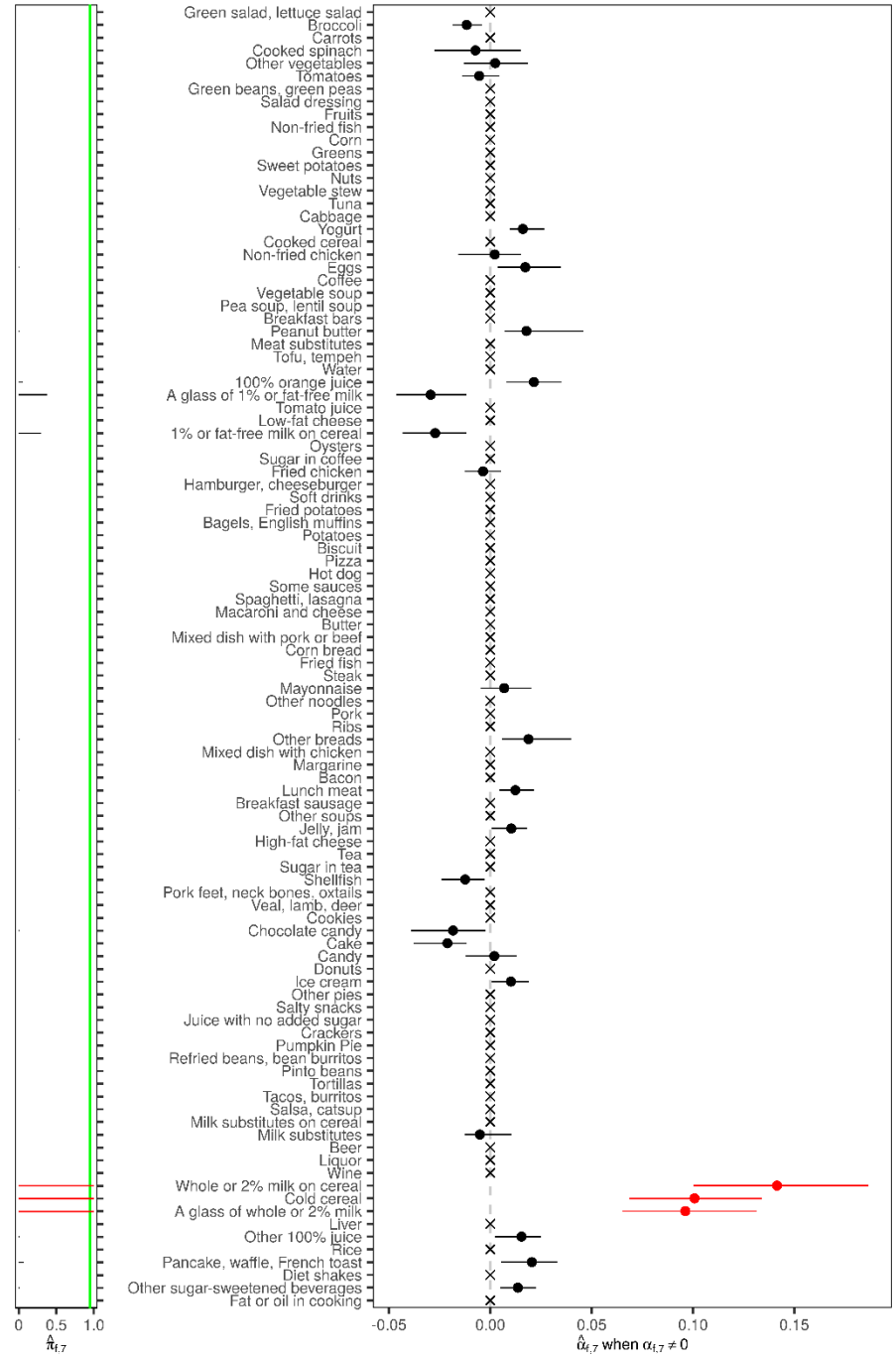

D
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,4}$
0.0
0.1
0.2
$\hat{\alpha}_{f,4}$ when $\alpha_{f,4} \neq 0$


E
Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,5}$
0.0
0.1
0.2
0.3
$\hat{\alpha}_{f,5}$ when $\alpha_{f,5} \neq 0$

F


Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{f,6}$
0.0
0.1
0.2
$\hat{\alpha}_{f,6}$ when $\alpha_{f,6} \neq 0$

G


Green salad, lettuce salad
Broccoli
Carrots
Cooked spinach
Other vegetables
Tomatoes
Green beans, green peas
Salad dressing
Fruits
Non-fried fish
Corn
Greens
Sweet potatoes
Nuts
Vegetable stew
Tuna
Cabbage
Yogurt
Cooked cereal
Non-fried chicken
Eggs
Coffee
Vegetable soup
Pea soup, lentil soup
Breakfast bars
Peanut butter
Meat substitutes
Tofu, tempeh
Water
100% orange juice
A glass of 1% or fat-free milk
Tomato juice
Low-fat cheese
1% or fat-free milk on cereal
Oysters
Sugar in coffee
Fried chicken
Hamburger, cheeseburger
Soft drinks
Fried potatoes
Bagels, English muffins
Potatoes
Biscuit
Pizza
Hot dog
Some sauces
Spaghetti, lasagna
Macaroni and cheese
Butter
Mixed dish with pork or beef
Corn bread
Fried fish
Steak
Mayonnaise
Other noodles
Pork
Ribs
Other breads
Mixed dish with chicken
Margarine
Bacon
Lunch meat
Breakfast sausage
Other soups
Jelly, jam
High-fat cheese
Tea
Sugar in tea
Shellfish
Pork feet, neck bones, oxtails
Veal, lamb, deer
Cookies
Chocolate candy
Cake
Candy
Donuts
Ice cream
Other pies
Salty snacks
Juice with no added sugar
Crackers
Pumpkin Pie
Refried beans, bean burritos
Pinto beans
Tortillas
Tacos, burritos
Salsa, catsup
Milk substitutes on cereal
Milk substitutes
Beer
Liquor
Wine
Whole or 2% milk on cereal
Cold cereal
A glass of whole or 2% milk
Liver
Other 100% juice
Rice
Pancake, waffle, French toast
Diet shakes
Other sugar-sweetened beverages
Fat or oil in cooking
0
0.5
1.0
$\hat{\pi}_{t,7}$
-0.05
0.00
0.05
0.10
0.15
$\hat{\alpha}_{t,7}$ when $\alpha_{t,7} \neq 0$

Posterior inclusion probability and factor loading estimate with its 95% credible interval. *Left panel* represents the posterior inclusion probability $\hat{\pi}_{f,j}$ that the corresponding factor loading is non-zero for food f and j-th dietary pattern. A green vertical line represents $\hat{\pi}_{f,j} = 0.95$. *Right panel* represents the posterior mean of factor loading $\hat{\alpha}_{i,j}$ with its 95% credible interval, summarizing where $\alpha_{i,j}$ is likely to be when $\alpha_{f,j} \neq 0$. $\hat{\pi}_{f,j}$ and $\hat{\alpha}_{i,j}$ allow foods that do not contribute to a dietary pattern to be distinguished from foods that have a relatively small but important contribution to a dietary pattern. 5000 Markov Chain Monte Carlo samples were used to estimate the model parameters. $\hat{\alpha}_{i,j}$ with its corresponding $\hat{\pi}_{f,j} < 0.001$ are considered as a point mass distribution at 0, and the cross symbol (×) was placed.

APPENDIX C

Model specifications for the estimation of factor scores

Suppose X_i is the $p \times t$ matrix that represents the measurements of dietary intake with p food items for individual i, collected at time points t ($t = 0$: baseline, $t = 1$: follow-up). Consider a factor-analytic dietary pattern model with k dietary patterns for a total of I individuals

$$X_i = AF_i + \upsilon_i = \sum_{j=1}^{K} A_j F_{ji}^T + \upsilon_i$$

where A is the $p \times k$ matrix of factor loadings with it columns A_j; F_i is the $k \times t$ factor score (dietary pattern score) for individual i; F_{ji}^T is the row vector of time-varying scores for individual i and j-th dietary pattern; υ_i is the $p \times t$ matrix of additive noise. Here, a factor loading matrix A represent how each dietary pattern is associated with observed intakes of p food and is assumed to be fixed in time. This ensures that the interpretation of dietary patterns does not change over time, and factor scores at different times are comparable on the same basis. Factor score F_i, on the other hand, is assumed to vary over time. To model factor score, let $F_{ji}(t)$ is the j-th dietary pattern score for individual i at time t. We then consider a multivariate linear mixed-effects model

$$F_{ji}(t) = \left(H_i \beta_j\right)(t) + \left(S_i \alpha_{ji}\right)(t) + \left(Z_i \gamma_{jc}\right)(t) + \epsilon_{ji}(t)$$

where H_i is the $t \times d$ matrix of linear predictors such as exercise characteristics and/or demographic variables for individual i; β_j is the d-dimensional vector of regression coefficients shared across all individuals; $\epsilon_{ji}(t)$ is the observation level errors and assumed to be normal. In our specific application, S_i and Z_i are the model matrices for random effects while α_{ji} and γ_{jc} are the corresponding random effects for individual i and cohort c, to account for the individual-dependent and cohort-dependent variations not explained by fixed effects β_j, respectively. The full model is then described as

$$X_{fi}(t) = \sum_{j=1}^{k} A_{fj}\left[\left(H_i\beta_j\right)(t) + \left(S_i\alpha_{ji}\right)(t) + \left(Z_i\gamma_{jc}\right)(t) + \epsilon_{ji}(t)\right] + \upsilon_{fi}(t)$$

where $X_{fi}(t)$ is the intake of food $f \in 1{:}\,p$ for individual i observed at time t; A_{fj} is the factor loading for food f in the j-th dietary pattern. For the Bayesian inference, we placed the following priors:

$$\upsilon_{fi} \sim N\left(0, \sigma_f\right)$$
$${\sigma_f}^{-1} \sim \text{gamma(a, b)}$$
$$\alpha_{ji} \sim N(0, \Sigma_{\alpha_j})$$
$$\gamma_{ji} \sim N(0, \Sigma_{\gamma_j})$$
$$\beta_{ji} \sim N(0, \Sigma_{\beta_j})$$

The covariance matricies, Σ_{α_j}, Σ_{γ_j}, and Σ_{β_j}, were decomposed with a correlation matrix and a vector of coefficient scales.

$$\Sigma = \text{diag\_matrix}(\tau)\ \Omega\ \text{diag\_matrix}(\tau)$$

For the componenet of scale vector τ, we placed an inverse gamma prior

$$\tau_k{}^{-1} \sim \text{gamma}(\text{c}, \text{d})$$

and gave the correlation matrix Ω an LKJ prior with shape $\upsilon \geq 1$,

$$\Omega \sim \text{LKJCorr}(\upsilon)$$

References

A Series of Systematic Reviews on the Relationship Between Dietary Patterns and Health Outcomes. (2014). Alexandria (VA): U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, Evidence Analysis Library Division. Retrieved from https://www.cnpp.usda.gov/sites/default/files/usda_nutrition_evidence_flbrary/DietaryPatternsExecutiveSummary.pdf

Aguilar, O., & West, M. (2000). Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business & Economic Statistics*, *18*(3), 338–357. https://doi.org/10.2307/1392266

Anderson, D. A., Shapiro, J. R., & Lundgren, J. D. (2003). The freshman year of college as a critical period for weight gain: An initial evaluation. *Eating Behaviors*, *4*(4), 363–367. https://doi.org/10.1016/S1471-0153(03)00030-8

Anderson, J. W., Konz, E. C., Frederich, R. C., & Wood, C. L. (2001). Long-term weight-loss maintenance: a meta-analysis of US studies. *The American Journal of Clinical Nutrition*, *74*(5), 579–584.

Ax, E., Warensjö Lemming, E., Becker, W., Andersson, A., Lindroos, A. K., Cederholm, T., … Fung, T. T. (2016). Dietary patterns in Swedish adults;

results from a national dietary survey. *British Journal of Nutrition*, *115*(01), 95–104. https://doi.org/10.1017/S0007114515004110

Baker, A. H., & Wardle, J. (2003). Sex differences in fruit and vegetable intake in older adults. *Appetite*, *40*(3), 269–275. https://doi.org/10.1016/S0195-6663(03)00014-X

Balder, H. F., Virtanen, M., Brants, H. A. M., Krogh, V., Dixon, L. B., Tan, F., … Goldbohm, R. A. (2003). Common and Country-Specific Dietary Patterns in Four European Cohort Studies. *The Journal of Nutrition*, *133*(12), 4246–4251.

Bales, C. W., Hawk, V. H., Granville, E. O., Rose, S. B., Shields, T., Bateman, L., … Kraus, W. E. (2012). Aerobic and Resistance Training Effects on Energy Intake: The STRRIDE AT/RT Study. *Medicine and Science in Sports and Exercise*, *44*(10), 2033–2039. https://doi.org/10.1249/MSS.0b013e318259479a

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2014). Demonstrating the validity of twin research in criminology. *Criminology*, *52*(4), 588–626.

Batres-Marquez, S. P., Jensen, H. H., & Upton, J. (2009). Rice Consumption in the United States: Recent Evidence from Food Consumption Surveys. *Journal of the American Dietetic Association*, *109*(10), 1719–1727. https://doi.org/10.1016/j.jada.2009.07.010

Berthoud, H.-R., Sutton, G. M., Townsend, R. L., Patterson, L. M., & Zheng, H. (2006). Brainstem mechanisms integrating gut-derived satiety signals and descending forebrain information in the control of meal size. *Physiology & Behavior*, *89*(4), 517–524. https://doi.org/10.1016/j.physbeh.2006.08.018

Blakely, F., Dunnagan, T., Haynes, G., Moore, S., & Pelican, S. (2004). Moderate Physical Activity and Its Relationship to Select Measures of a Healthy Diet. *The Journal of Rural Health*, *20*(2), 160–165. https://doi.org/10.1111/j.1748-0361.2004.tb00023.x

Blundell, J. E., Gibbons, C., Caudwell, P., Finlayson, G., & Hopkins, M. (2015). Appetite control and energy balance: impact of exercise. *Obesity Reviews*, *16*, 67–76. https://doi.org/10.1111/obr.12257

Blundell, J. E., Stubbs, R. J., Hughes, D. A., Whybrow, S., & King, N. A. (2003). Cross talk between physical activity and appetite control: does physical activity stimulate appetite? *The Proceedings of the Nutrition Society*, *62*(3), 651–661. https://doi.org/10.1079/PNS2003286

Brandon, L. J., & Elliott-Lloyd, M. B. (2006). Walking, body composition, and blood pressure dose-response in African American and white women. *Ethnicity & Disease*, *16*(3), 675–681.

Bree, M. B. van den, Eaves, L. J., & Dwyer, J. T. (1999). Genetic and environmental influences on eating patterns of twins aged ≥50 y. *The American Journal of Clinical Nutrition*, *70*(4), 456–465.

Breen, F. M., Plomin, R., & Wardle, J. (2006). Heritability of food preferences in young children. *Physiology & Behavior*, *88*(4–5), 443–447. https://doi.org/10.1016/j.physbeh.2006.04.016

Breslin, P. A. S. (2013). An Evolutionary Perspective on Food and Human Taste. *Current Biology*, *23*(9), R409–R418. https://doi.org/10.1016/j.cub.2013.04.010

Broom, D. R., Stensel, D. J., Bishop, N. C., Burns, S. F., & Miyashita, M. (2007). Exercise-induced suppression of acylated ghrelin in humans. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *102*(6), 2165–2171. https://doi.org/10.1152/japplphysiol.00759.2006

Browne, M. W. (2001). An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Brunt, A. R., & Rhee, Y. S. (2008). Obesity and lifestyle in U.S. college students related to living arrangemeents. *Appetite*, *51*(3), 615–621. https://doi.org/10.1016/j.appet.2008.04.019

Cadima, J. F., & Jolliffe, I. T. (2001). Variable selection and the interpretation of principal subspaces. *Journal of Agricultural, Biological, and Environmental Statistics*, *6*(1), 62–79.

Campos, G. de los, Vazquez, A. I., Fernando, R., Klimentidis, Y. C., & Sorensen, D. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLOS Genetics*, *9*(7), e1003608. https://doi.org/10.1371/journal.pgen.1003608

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., & West, M. (2008). High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, *103*(484), 1438–1456. https://doi.org/10.1198/016214508000000869

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling Sparsity via the Horseshoe, 8.

Cattell, Raymond B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cattell, Raymond Bernard. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Plenum Press.

Cecil, J. E., Tavendale, R., Watt, P., Hetherington, M. M., & Palmer, C. N. A. (2008). An Obesity-Associated FTO Gene Variant and Increased Energy Intake in Children. *New England Journal of Medicine*, *359*(24), 2558–2566. https://doi.org/10.1056/NEJMoa0803839

Chamoun, E., Mutch, D. M., Allen-Vercoe, E., Buchholz, A. C., Duncan, A. M., Spriet, L. L., … on behalf of the Guelph Family Health Study. (2018). A review of the associations between single nucleotide polymorphisms in taste receptors, eating behaviors, and health. *Critical Reviews in Food Science and Nutrition*, *58*(2), 194–207. https://doi.org/10.1080/10408398.2016.1152229

Charlton, K., Kowal, P., Soriano, M. M., Williams, S., Banks, E., Vo, K., & Byles, J. (2014). Fruit and Vegetable Intake and Body Mass Index in a Large Sample of Middle-Aged Australian Men and Women. *Nutrients*, *6*(6), 2305–2319. https://doi.org/10.3390/nu6062305

Chen, W., Wang, H. J., Shang, N. N., Liu, J., Li, J., Tang, D. H., & Li, Q. (2017). Moderate intensity treadmill exercise alters food preference via dopaminergic plasticity of ventral tegmental area-nucleus accumbens in obese mice. *Neuroscience Letters*, *641*, 56–61. https://doi.org/10.1016/j.neulet.2017.01.055

Chmurzynska, A., & Mlodzik, M. A. (2017). Genetics of fat intake in the determination of body mass. *Nutrition Research Reviews*, *30*(01), 106–117. https://doi.org/10.1017/S0954422417000014

Clarke, T.-K., Adams, M., Davies, G., Howard, D., Hall, L., Padmanabhan, S., … McIntosh, A. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=122 117). *Molecular Psychiatry*, 9.

Clément, K., Vaisse, C., Lahlou, N., Cabrol, S., Pelloux, V., Cassuto, D., … Guy-Grand, B. (1998). A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*, *392*(6674), 398–401. https://doi.org/10.1038/32911

Craft, B. B., Carroll, H. A., & Lustyk, M. K. B. (2014). Gender Differences in Exercise Habits and Quality of Life Reports: Assessing the Moderating Effects of Reasons for Exercise. *International Journal of Liberal Arts and Social Science*, *2*(5), 65–76.

Cutler, G. J., Flood, A., Hannan, P. J., Slavin, J. L., & Neumark-Sztainer, D. (2012). Association between major patterns of dietary intake and weight status in adolescents. *British Journal of Nutrition*, *108*(02), 349–356. https://doi.org/10.1017/S0007114511005435

de Castro, J. M. (1993). Genetic influences on daily intake and meal patterns of humans. *Physiology & Behavior*, *53*(4), 777–782. https://doi.org/10.1016/0031-9384(93)90188-L

de los Campos, G., Gianola, D., & Allison, D. B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nature Reviews Genetics*, *11*(12), 880–886. https://doi.org/10.1038/nrg2898

Dekker, L. H., Dam, R. M. van, Snijder, M. B., Peters, R. J., Dekker, J. M., Vries, J. H. de, … Nicolaou, M. (2015). Comparable Dietary Patterns Describe Dietary Behavior across Ethnic Groups in the Netherlands, but Different Elements in the Diet Are Associated with Glycated Hemoglobin and Fasting Glucose Concentrations. *The Journal of Nutrition*, *145*(8), 1884–1891. https://doi.org/10.3945/jn.114.207472

Deliens, T., Clarys, P., De Bourdeaudhuij, I., & Deforche, B. (2014). Determinants of eating behaviour in university students: a qualitative study using focus group discussions. *BMC Public Health*, *14*, 53. https://doi.org/10.1186/1471-2458-14-53

den Hoed, M., Westerterp-Plantenga, M. S., Bouwman, F. G., Mariman, E. C., & Westerterp, K. R. (2009). Postprandial responses in hunger and satiety are associated with the rs9939609 single nucleotide polymorphism in FTO. *The American Journal of Clinical Nutrition*, *90*(5), 1426–1432. https://doi.org/10.3945/ajcn.2009.28053

Donnelly, J. E., Herrmann, S. D., Lambourne, K., Szabo, A. N., Honas, J. J., & Washburn, R. A. (2014). Does Increased Exercise or Physical Activity Alter Ad-Libitum Daily Energy Intake or Macronutrient Composition in Healthy Adults? A Systematic Review. *PLoS ONE*, *9*(1), e83498. https://doi.org/10.1371/journal.pone.0083498

Drewnowski, A., & Darmon, N. (2005). The economics of obesity: dietary energy density and energy cost. *The American Journal of Clinical Nutrition*, *82*(1), 265S-273S. https://doi.org/10.1093/ajcn/82.1.265S

Dubowitz, T., Heron, M., Bird, C. E., Lurie, N., Finch, B. K., Basurto-Dávila, R., … Escarce, J. J. (2008). Neighborhood socioeconomic status and fruit and vegetable intake among Whites, Blacks, and Mexican-Americans in the United States. *The American Journal of Clinical Nutrition*, *87*(6), 1883–1891.

Dušátková, L., Zamrazilová, H., Aldhoon-Hainerová, I., Sedláčková, B., Včelák, J., Hlavatý, P., … Hainer, V. (2015). A common variant near BDNF is associated with dietary calcium intake in adolescents. *Nutrition Research*, *35*(9), 766–773. https://doi.org/10.1016/j.nutres.2015.06.004

Eaton, C. B., McPhillips, J. B., Gans, K. M., Garber, C. E., Assaf, A. R., Lasater, T. M., & Carleton, R. A. (1995). Cross-sectional relationship between diet and physical activity in two southeastern New England communities.

American Journal of Preventive Medicine. Retrieved from http://psycnet.apa.org/psycinfo/1996-12483-001

Elder, S. J., & Roberts, S. B. (2007). The Effects of Exercise on Food Intake and Body Fatness: A Summary of Published Studies. *Nutrition Reviews*, *65*(1), 1–19. https://doi.org/10.1301/nr.2007.jan.1-19

Emmons, K. M., McBride, C. M., Puleo, E., Pollak, K. I., Marcus, B. H., Napolitano, M., … Fletcher, R. (2005). Prevalence and predictors of multiple behavioral risk factors for colon cancer. *Preventive Medicine*, *40*(5), 527–534. https://doi.org/10.1016/j.ypmed.2004.10.001

Eny, K. M., Wolever, T. M., Corey, P. N., & El-Sohemy, A. (2010). Genetic variation in TAS1R2 (Ile191Val) is associated with consumption of sugars in overweight and obese individuals in 2 distinct populations. *The American Journal of Clinical Nutrition*, *92*(6), 1501–1510. https://doi.org/10.3945/ajcn.2010.29836

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fahey, M. T., Thane, C. W., Bramwell, G. D., & Coward, W. A. (2007). Conditional Gaussian mixture modelling for dietary pattern analysis.

Journal of the Royal Statistical Society: Series A (Statistics in Society), *170*(1), 149–166. https://doi.org/10.1111/j.1467-985X.2006.00452.x

Faith, M. S., Rhea, S. A., Corley, R. P., & Hewitt, J. K. (2008). Genetic and shared environmental influences on children's 24-h food and beverage intake: sex differences at age 7 y. *The American Journal of Clinical Nutrition*, *87*(4), 903–911.

Farah, N. M. F., Malkova, D., & Gill, J. M. R. (2010). Effects of Exercise on Postprandial Responses to Ad Libitum Feeding in Overweight Men: *Medicine & Science in Sports & Exercise*, *42*(11), 2015–2022. https://doi.org/10.1249/MSS.0b013e3181e0d186

Farooqi, I. S., & O'Rahilly, S. (2006). Genetics of Obesity in Humans. *Endocrine Reviews*, *27*(7), 710–718. https://doi.org/10.1210/er.2006-0040

Fatima, W., Shahid, A., Imran, M., Manzoor, J., Hasnain, S., Rana, S., & Mahmood, S. (2011). Leptin deficiency and leptin gene mutations in obese children from Pakistan. *International Journal of Pediatric Obesity: IJPO: An Official Journal of the International Association for the Study of Obesity*, *6*(5–6), 419–427. https://doi.org/10.3109/17477166.2011.608431

Fawcett, K. A., & Barroso, I. (2010). The genetics of obesity: FTO leads the way. *Trends in Genetics*, *26*(6), 266–274. https://doi.org/10.1016/j.tig.2010.02.006

Fedewa, M. V., Das, B. M., Evans, E. M., & Dishman, R. K. (2014). Change in Weight and Adiposity in College Students. *American Journal of Preventive Medicine*, *47*(5), 641–652. https://doi.org/10.1016/j.amepre.2014.07.035

Field, A. E., Gillman, M. W., Rosner, B., Rockett, H. R., & Colditz, G. A. (2003). Association between fruit and vegetable intake and change in body mass index among a large sample of children and adolescents in the United States. *International Journal of Obesity*, *27*(7), 821–826. https://doi.org/10.1038/sj.ijo.0802297

Fildes, A., Jaarsveld, V., Hm, C., Llewellyn, C. H., Fisher, A., Cooke, L., & Wardle, J. (2014). Nature and nurture in children's food preferences. *The American Journal of Clinical Nutrition*, *99*(4), 911–917. https://doi.org/10.3945/ajcn.113.077867

Flood, A., Rastogi, T., Wirfält, E., Mitrou, P. N., Reedy, J., Subar, A. F., … others. (2008). Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged Americans. *The American Journal of Clinical Nutrition*, *88*(1), 176–184.

Franko, D. L., Cousineau, T. M., Trant, M., Green, T. C., Rancourt, D., Thompson, D., … Ciccazzo, M. (2008). Motivation, self-efficacy, physical activity and nutrition in college students: Randomized controlled trial of

an Internet-based education program. *Preventive Medicine*, *47*(4), 369–377. https://doi.org/10.1016/j.ypmed.2008.06.013

Fraser, G. E., Welch, A., Luben, R., Bingham, S. A., & Day, N. E. (2000). The Effect of Age, Sex, and Education on Food Consumption of a Middle-Aged English Cohort—EPIC in East Anglia. *Preventive Medicine*, *30*(1), 26–34. https://doi.org/10.1006/pmed.1999.0598

Gallagher, M. L., Farrior, E., Broadhead, L., Gillette, L. S., Rowe, M. L., Somes, G., … Kolasa, K. M. (1993). Development and testing of a food frequency recall instrument for describing dietary patterns in adults and teenagers. *Nutrition Research*, *13*(2), 177–188.

Gao, C., Brown, C. D., & Engelhardt, B. E. (2013). A latent factor model with a mixture of sparse and dense factors to model gene expression data with confounding effects. *ArXiv Preprint ArXiv:1310.4792*.

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., … Farrer, L. A. (2014). Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Molecular Psychiatry*, *19*(1), 41–49. https://doi.org/10.1038/mp.2013.145

George, E. I., & Clyde, M. (2004). Model Uncertainty. *Statistical Science*, *19*(1), 81–94. https://doi.org/10.1214/088342304000000035

Geweke, J. (1993). Bayesian Treatment of the Independent Student-t Linear Model. *Journal of Applied Econometrics*, *8*, S19–S40.

Geweke, John, & Zhou, G. (1996). Measuring the Pricing Error of the Arbitrage Pricing Theory. *The Review of Financial Studies*, *9*(2), 557–587. https://doi.org/10.1093/rfs/9.2.557

Ghahramani, Z., & Griffiths, T. L. (2005). Infinite latent feature models and the Indian buffet process. In *Advances in neural information processing systems* (pp. 475–482). Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2005_130.pdf

Gibson, G. (2010). Hints of hidden heritability in GWAS. *Nature Genetics; New York*, *42*(7), 558–560.

Gordon-Larsen, P., Nelson, M. C., & Popkin, B. M. (2004). Longitudinal physical activity and sedentary behavior trends: Adolescence to adulthood. *American Journal of Preventive Medicine*, *27*(4), 277–283. https://doi.org/10.1016/j.amepre.2004.07.006

Gorst-Rasmussen, A., Dahm, C. C., Dethlefsen, C., Scheike, T., & Overvad, K. (2011). Exploring Dietary Patterns By Using the Treelet Transform. *American Journal of Epidemiology*, *173*(10), 1097–1104. https://doi.org/10.1093/aje/kwr060

Grimm, E. R., & Steinle, N. I. (2011). Genetics of eating behavior: established and emerging concepts. *Nutrition Reviews*, *69*(1), 52–60. https://doi.org/10.1111/j.1753-4887.2010.00361.x

Gropper, S. S., Simmons, K. P., Connell, L. J., & Ulrich, P. V. (2012). Changes in body weight, composition, and shape: a 4-year study of college students. *Applied Physiology, Nutrition, and Metabolism*, *37*(6), 1118–1123. https://doi.org/10.1139/h2012-139

Guenther, P. M., Casavale, K. O., Kirkpatrick, S. I., Reedy, J., Hiza, H. A. B., Kuczynski, K. J., … Krebs-Smith, S. M. (2013). Update of the Healthy Eating Index: HEI-2010. *Journal of the Academy of Nutrition and Dietetics*, *113*(4). https://doi.org/10.1016/j.jand.2012.12.016

Gunderson, E. P., Tsai, A.-L., Selby, J. V., Caan, B., Mayer-Davis, E. J., & Risch, N. (2006). Twins of Mistaken Zygosity (TOMZ): Evidence for Genetic Contributions to Dietary Patterns and Physiologic Traits. *Twin Research and Human Genetics*, *9*(4), 540–549. https://doi.org/10.1375/183242706778025053

Hall, K. D., Heymsfield, S. B., Kemnitz, J. W., Klein, S., Schoeller, D. A., & Speakman, J. R. (2012). Energy balance and its components: implications for body weight regulation123. *The American Journal of Clinical Nutrition*, *95*(4), 989–994. https://doi.org/10.3945/ajcn.112.036350

Handa, K., & Kreiger, N. (2002). Diet patterns and the risk of renal cell carcinoma. *Public Health Nutrition*, *5*(06), 757–767. https://doi.org/10.1079/PHN2002347

Hasselbalch, A. L., Heitmann, B. L., Kyvik, K. O., & Sørensen, T. I. A. (2008). Studies of Twins Indicate That Genetics Influence Dietary Intake. *The Journal of Nutrition*, *138*(12), 2406–2412. https://doi.org/10.3945/jn.108.087668

Herring, M. P., Sailors, M. H., & Bray, M. S. (2014). Genetic factors in exercise adoption, adherence and obesity. *Obesity Reviews*, *15*(1), 29–39. https://doi.org/10.1111/obr.12089

Hess, J. M., Jonnalagadda, S. S., & Slavin, J. L. (2016). What Is a Snack, Why Do We Snack, and How Can We Choose Better Snacks? A Review of the Definitions of Snacking, Motivations to Snack, Contributions to Dietary Intake, and Recommendations for Improvement. *Advances in Nutrition: An International Review Journal*, *7*(3), 466–475. https://doi.org/10.3945/an.115.009571

Hinney, A., Hoch, A., Geller, F., Schäfer, H., Siegfried, W., Goldschmidt, H., … Hebebrand, J. (2002). Ghrelin gene: identification of missense variants and a frameshift mutation in extremely obese children and adolescents and healthy normal weight students. *The Journal of Clinical Endocrinology and Metabolism*, *87*(6), 2716. https://doi.org/10.1210/jcem.87.6.8672

Hoffmann, K., Schulze, M. B., Schienkiewitz, A., Nöthlings, U., & Boeing, H. (2004). Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *American Journal of Epidemiology*, *159*(10), 935–944. https://doi.org/10.1093/aje/kwh134

Horio, T. (2004). Effect of physical exercise on human preference for solutions of various sweet substances. *Perceptual and Motor Skills*, *99*(3), 1061–1070.

Horio, T., & Kawamura, Y. (1998). Influence of physical exercise on human preferences for various taste solutions. *Chemical Senses*, *23*(4), 417–421.

Horsch, A., Wobmann, M., Kriemler, S., Munsch, S., Borloz, S., Balz, A., … Puder, J. J. (2015). Impact of physical activity on energy balance, food intake and choice in normal weight and obese children in the setting of acute social stress: a randomized controlled trial. *BMC Pediatrics*, *15*(1), 12. https://doi.org/10.1186/s12887-015-0326-7

Hu, F. B. (2002). Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology*, *13*(1), 3–9.

Hu, F. B. (2008). *Obesity epidemiology*. Oxford ; New York: Oxford University Press.

Hu, F. B., Stampfer, M. J., Rimm, E., Ascherio, A., Rosner, B. A., Spiegelman, D., & Willett, W. C. (1999). Dietary Fat and Coronary Heart Disease: A Comparison of Approaches for Adjusting for Total Energy Intake and

Modeling Repeated Dietary Measurements. *American Journal of Epidemiology*, *149*(6), 531–540.

Hur, Y.-M., Bouchard, T. J., & Eckert, E. (1998). Genetic and environmental influences on self-reported diet: a reared-apart twin study. *Physiology & Behavior*, *64*(5), 629–636.

Huvenne, H., & Dubern, B. (2014). Monogenic Forms of Obesity. In C. Nóbrega & R. Rodriguez-López (Eds.), *Molecular Mechanisms Underpinning the Development of Obesity* (pp. 9–21). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12766-8_2

Imamura, F., & Jacques, P. F. (2011). Invited Commentary: Dietary Pattern Analysis. *American Journal of Epidemiology*, *173*(10), 1105–1108. https://doi.org/10.1093/aje/kwr063

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. https://doi.org/10.1007/BF02289233

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kanarek, R. B., Ryu, M., & Przypek, J. (1995). Preferences for foods with varying levels of salt and fat differ as a function of dietary restraint and exercise but not menstrual cycle. *Physiology & Behavior*, *57*(5), 821–826.

Kano, Y., Miyamoto, Y., & Shimizu, S. (2003). Factor rotation and ICA. In *Proceedings of Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)* (Vol. 101, p. 105). Retrieved from http://www.kecl.ntt.co.jp/icl/signal/ica2003/cdrom/data/0155.pdf

Kattelmann, K. K., White, A. A., Greene, G. W., Byrd-Bredbenner, C., Hoerr, S. L., Horacek, T. M., … Morrell, J. S. (2014). Development of Young Adults Eating and Active for Health (YEAH) Internet-Based Intervention via a Community-Based Participatory Research Model. *Journal of Nutrition Education and Behavior*, *46*(2), S10–S25. https://doi.org/10.1016/j.jneb.2013.11.006

Keller, K. L., Pietrobelli, A., Must, S., & Faith, M. S. (2002). Genetics of eating and its relation to obesity. *Current Atherosclerosis Reports*, *4*(3), 176–182.

Keskitalo, K., Silventoinen, K., Tuorila, H., Perola, M., Pietiläinen, K. H., Rissanen, A., & Kaprio, J. (2008). Genetic and environmental contributions to food use patterns of young adult twins. *Physiology & Behavior*, *93*(0), 235–242. https://doi.org/10.1016/j.physbeh.2007.08.025

Kim, U. -k. (2003). Positional Cloning of the Human Quantitative Trait Locus Underlying Taste Sensitivity to Phenylthiocarbamide. *Science*, *299*(5610), 1221–1225. https://doi.org/10.1126/science.1080190

King, N. A., Burley, V. J., Blundell, J. E., & others. (1994). Exercise-induced suppression of appetite: effects on food intake and implications for energy balance. *European Journal of Clinical Nutrition*, *48*(10), 715–724.

King, N. A., Hopkins, M., Caudwell, P., Stubbs, R. J., & Blundell, J. E. (2009). Beneficial effects of exercise: shifting the focus from body weight to other markers of health. *British Journal of Sports Medicine*, *43*(12), 924–927. https://doi.org/10.1136/bjsm.2009.065557

King, N. A., Lluch, A., Stubbs, R. J., & Blundell, J. E. (1997). High dose exercise does not increase hunger or energy intake in free living males. *European Journal of Clinical Nutrition*, *51*(7), 478–483.

Kirkwood, L., Aldujaili, E., & Drummond, S. (2007). Effects of advice on dietary intake and/or physical activity on body composition, blood lipids and insulin resistance following a low-fat, sucrose-containing, high-carbohydrate, energy-restricted diet. *International Journal of Food Sciences and Nutrition*, *58*(5), 383–397. https://doi.org/10.1080/09637480701252336

Knäuper, B., Rabiau, M., Cohen, O., & Patriciu, N. (2004). Compensatory health beliefs: scale development and psychometric properties. *Psychology & Health*, *19*(5), 607–624. https://doi.org/10.1080/0887044042000196737

Knowles, D., & Ghahramani, Z. (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *The Annals of*

Applied Statistics, *5*(2B), 1534–1552. https://doi.org/10.1214/10-AOAS435

Kolodinsky, J., Harvey-Berino, J. R., Berlin, L., Johnson, R. K., & Reynolds, T. W. (2007). Knowledge of Current Dietary Guidelines and Food Choice by College Students: Better Eaters Have Higher Knowledge of Dietary Guidance. *Journal of the American Dietetic Association*, *107*(8), 1409–1413. https://doi.org/10.1016/j.jada.2007.05.016

Korbonits, M., Gueorguiev, M., O'Grady, E., Lecoeur, C., Swan, D. C., Mein, C. A., … Froguel, P. (2002). A variation in the ghrelin gene increases weight and decreases insulin secretion in tall, obese children. *The Journal of Clinical Endocrinology and Metabolism*, *87*(8), 4005–4008. https://doi.org/10.1210/jcem.87.8.8881

Krom, M. de, Schouw, Y. T. van der, Hendriks, J., Ophoff, R. A., Gils, C. H. van, Stolk, R. P., … Adan, R. (2007). Common Genetic Variations in CCK, Leptin, and Leptin Receptor Genes Are Associated With Specific Human Eating Patterns. *Diabetes*, *56*(1), 276–280. https://doi.org/10.2337/db06-0473

Krude, H., Biebermann, H., Luck, W., Horn, R., Brabant, G., & Grüters, A. (1998). Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genetics*, *19*(2), 155–157. https://doi.org/10.1038/509

Laskowski, E. R. (2012). The Role of Exercise in the Treatment of Obesity. *PM&R*, *4*(11), 840–844. https://doi.org/10.1016/j.pmrj.2012.09.576

Lee, J. R., Muckerman, J. E., Wright, A. M., Davis, D. J., Childs, T. E., Gillespie, C. E., … Will, M. J. (2017). Sex determines effect of physical activity on diet preference: Association of striatal opioids and gut microbiota composition. *Behavioural Brain Research*, *334*, 16–25. https://doi.org/10.1016/j.bbr.2017.07.018

Leshem, M., Abutbul, A., & Eilon, R. (1999). Exercise increases the preference for salt in humans. *Appetite*, *32*(2), 251–260.

Li, J., Liu, H., Beaty, T. H., Chen, H., Caballero, B., & Wang, Y. (2016). Heritability of Children's Dietary Intakes: A Population-Based Twin Study in China. *Twin Research and Human Genetics*, *19*(05), 472–484. https://doi.org/10.1017/thg.2016.61

Liang, N.-C., Bello, N. T., & Moran, T. H. (2015). Wheel running reduces high-fat diet intake, preference and mu-opioid agonist stimulated intake. *Behavioural Brain Research*, *284*, 1–10. https://doi.org/10.1016/j.bbr.2015.02.004

Liebman, M., Propst, K., Moore, S. A., Pelican, S., Holmes, B., Wardlaw, M. K., … Dunnagan, T. (2003). Gender differences in selected dietary intakes and eating behaviors in rural communities in Wyoming, Montana, and

Idaho. *Nutrition Research*, *23*(8), 991–1002. https://doi.org/10.1016/S0271-5317(03)00080-0

Liu, J., Tuvblad, C., Raine, A., & Baker, L. (2013). Genetic and environmental influences on nutrient intake. *Genes & Nutrition*, *8*(2), 241–252. https://doi.org/10.1007/s12263-012-0320-8

Loos, R. J. F., Lindgren, C. M., Li, S., Wheeler, E., Zhao, J. H., Prokopenko, I., … Mohlke, K. L. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics*, *40*(6), 768–775. https://doi.org/10.1038/ng.140

Lucas, J., Carvalho, C., Wang, Q., Bild, A., Nevins, J. R., & West, M. (2006). Sparse statistical modelling in gene expression genomics. In *Bayesian Inference for Gene Expression and Proteomics* (pp. 155–176). Cambridge, United Kingdom, Cambridge University Press.

Lunetta, K. L. (2008). Genetic Association Studies. *Circulation*, *118*(1), 96–101. https://doi.org/10.1161/CIRCULATIONAHA.107.700401

Lutomski, J. E., van den Broeck, J., Harrington, J., Shiely, F., & Perry, I. J. (2011). Sociodemographic, lifestyle, mental health and dietary factors associated with direction of misreporting of energy intake. *Public Health Nutrition*, *14*(03), 532–541. https://doi.org/10.1017/S1368980010001801

Marquis, M. (2005). Exploring convenience orientation as a food motivation for college students living in residence halls. *International Journal of*

Consumer Studies, *29*(1), 55–63. https://doi.org/10.1111/j.1470-6431.2005.00375.x

Martinez, M. E., Marshall, J. R., & Sechrest, L. (1998). Invited commentary: factor analysis and the search for objectivity. *American Journal of Epidemiology*, *148*(1), 17–19.

Martins, C., Morgan, L., & Truby, H. (2008). A review of the effects of exercise on appetite regulation: an obesity perspective. *International Journal of Obesity*, *32*(9), 1337–1347. https://doi.org/10.1038/ijo.2008.98

Martins, Catia, Stensvold, D., Finlayson, G., Holst, J., Wisloff, U., Kulseng, B., … King, N. A. (2015). Effect of Moderate- and High-Intensity Acute Exercise on Appetite in Obese Individuals: *Medicine & Science in Sports & Exercise*, *47*(1), 40–48. https://doi.org/10.1249/MSS.0000000000000372

McCaffery, J. M., Jablonski, K. A., Franks, P. W., Delahanty, L. M., Aroda, V., Marrero, D., … Florez, J. C. (2017). Replication of the Association of BDNF and MC4R Variants With Dietary Intake in the Diabetes Prevention Program., Replication of the association of BDNF and MC4R variants with dietary intake in the Diabetes Prevention Program. *Psychosomatic Medicine, Psychosomatic Medicine*, *79*, *79*(2, 2), 224, 224–233. https://doi.org/10.1097/PSY.0000000000000380, 10.1097/PSY.0000000000000380

McCaffery, Jeanne M., Papandonatos, G. D., Peter, I., Huggins, G. S., Raynor, H. A., Delahanty, L. M., … Group, for T. G. S. of L. A. and T. L. A. R. (2012). Obesity susceptibility loci and dietary intake in the Look AHEAD Trial. *The American Journal of Clinical Nutrition*, *95*(6), 1477–1486. https://doi.org/10.3945/ajcn.111.026955

McGowan, L., Croker, H., Wardle, J., & Cooke, L. J. (2012). Environmental and individual determinants of core and non-core food and drink intake in preschool-aged children in the United Kingdom. *European Journal of Clinical Nutrition*, *66*(3), 322–328. https://doi.org/10.1038/ejcn.2011.224

McNeil, J., Cadieux, S., Finlayson, G., Blundell, J. E., & Doucet, É. (2015). The effects of a single bout of aerobic or resistance exercise on food reward. *Appetite*, *84*, 264–270. https://doi.org/10.1016/j.appet.2014.10.018

McTigue, K. M., Garrett, J. M., & Popkin, B. M. (2002). The natural history of the development of obesity in a cohort of young US adults between 1981 and 1998. *Annals of Internal Medicine*, *136*(12), 857–864.

Melhorn, S. J., Askren, M. K., Chung, W. K., Kratz, M., Bosch, T. A., Tyagi, V., … Schur, E. A. (2018). FTO genotype impacts food intake and corticolimbic activation. *The American Journal of Clinical Nutrition*, *107*(2), 145–154. https://doi.org/10.1093/ajcn/nqx029

Miller, F. L., O'Connor, D. P., Herring, M. P., Sailors, M. H., Jackson, A. S., Dishman, R. K., & Bray, M. S. (2014). Exercise Dose, Exercise

Adherence, and Associated Health Outcomes in the TIGER Study. *Medicine and Science in Sports and Exercise*, *46*(1), 69–75. https://doi.org/10.1249/MSS.0b013e3182a038b9

Moeller, S. M., Reedy, J., Millen, A. E., Dixon, L. B., Newby, P. K., Tucker, K. L., … Guenther, P. M. (2007). Dietary Patterns: Challenges and Opportunities in Dietary Patterns Research. *Journal of the American Dietetic Association*, *107*(7), 1233–1239. https://doi.org/10.1016/j.jada.2007.03.014

Montague, C. T., Farooqi, I. S., Whitehead, J. P., Soos, M. A., Rau, H., Wareham, N. J., … O'Rahilly, S. (1997). Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, *387*(6636), 903–908. https://doi.org/10.1038/43185

Monteleone, P., Tortorella, A., Castaldo, E., Di Filippo, C., & Maj, M. (2007). The Leu72Met polymorphism of the ghrelin gene is significantly associated with binge eating disorder: *Psychiatric Genetics*, *17*(1), 13–16. https://doi.org/10.1097/YPG.0b013e328010e2c3

Moody, L., Liang, J., Choi, P. P., Moran, T. H., & Liang, N.-C. (2015). Wheel running decreases palatable diet preference in Sprague–Dawley rats. *Physiology & Behavior*, *150*, 53–63. https://doi.org/10.1016/j.physbeh.2015.03.019

Nakazato, M., Murakami, N., Date, Y., Kojima, M., Matsuo, H., Kangawa, K., & Matsukura, S. (2001). A role for ghrelin in the central regulation of feeding. *Nature*, *409*(6817), 194–198. https://doi.org/10.1038/35051587

Nanri, A., Mizoue, T., Shimazu, T., Ishihara, J., Takachi, R., Noda, M., … Group, for the J. P. H. C.-B. P. S. (2017). Dietary patterns and all-cause, cancer, and cardiovascular disease mortality in Japanese men and women: The Japan public health center-based prospective study. *PLOS ONE*, *12*(4), e0174848. https://doi.org/10.1371/journal.pone.0174848

Neale, B. M., Mazzeo, S. E., & Bulik, C. M. (2003). A twin study of dietary restraint, disinhibition and hunger: an examination of the eating inventory (three factor eating questionnaire). *Twin Research: The Official Journal of the International Society for Twin Studies*, *6*(6), 471–478. https://doi.org/10.1375/136905203322686455

Nettleton, J. A., Steffen, L. M., Ni, H., Liu, K., & Jacobs, D. R. (2008). Dietary Patterns and Risk of Incident Type 2 Diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care*, *31*(9), 1777–1782. https://doi.org/10.2337/dc08-0760

Newby, P. K., & Tucker, K. L. (2004). Empirically Derived Eating Patterns Using Factor or Cluster Analysis: A Review. *Nutrition Reviews*, *62*(5), 177–203. https://doi.org/10.1111/j.1753-4887.2004.tb00040.x

Nigg, C. R., Burbank, P. M., Padula, C., Dufresne, R., Rossi, J. S., Velicer, W. F., … Prochaska, J. O. (1999). Stages of change across ten health risk behaviors for older adults. *The Gerontologist*, *39*(4), 473–482.

Northstone, K., Ness, A., Emmett, P., & Rogers, I. (2008). Adjusting for energy intake in dietary pattern investigations using principal components analysis. *European Journal of Clinical Nutrition*, *62*(7), 931–938. https://doi.org/10.1038/sj.ejcn.1602789

Oh, H., & Taylor, A. H. (2012). Brisk walking reduces ad libitum snacking in regular chocolate eaters during a workplace simulation. *Appetite*, *58*(1), 387–392. https://doi.org/10.1016/j.appet.2011.11.006

Oh, H., & Taylor, A. H. (2014). Self-regulating smoking and snacking through physical activity. *Health Psychology*, *33*(4), 349–359. https://doi.org/10.1037/a0032423

Oppert, J.-M., Thomas, F., Charles, M.-A., Benetos, A., Basdevant, A., & Simon, C. (2006). Leisure-time and occupational physical activity in relation to cardiovascular risk factors and eating habits in French adults. *Public Health Nutrition*, *9*(06). https://doi.org/10.1079/PHN2005882

O'rahilly, S., & Farooqi, I. S. (2008). Human obesity as a heritable disorder of the central control of energy balance. *International Journal of Obesity; London*, *32*(S7), S55-61. http://dx.doi.org/10.1038/ijo.2008.239

Oswal, A., & Yeo, G. S. H. (2007). The leptin melanocortin pathway and the control of body weight: lessons from human and murine genetics. *Obesity Reviews*, *8*(4), 293–306. https://doi.org/10.1111/j.1467-789X.2007.00378.x

Pallister, T., Spector, T. D., & Menni, C. (2014). Twin studies advance the understanding of gene–environment interplay in human nutrigenomics. *Nutrition Research Reviews*, *27*(02), 242–251. https://doi.org/10.1017/S095442241400016X

Panek, L. M., Jones, K. R., & Temple, J. L. (2014). Short term aerobic exercise alters the reinforcing value of food in inactive adults. *Appetite*, *81*, 320–329. https://doi.org/10.1016/j.appet.2014.06.102

Park, S. L., Cheng, I., Pendergrass, S. A., Kucharska-Newton, A. M., Lim, U., Ambite, J. L., … Le Marchand, L. (2013). Association of the FTO Obesity Risk Variant rs8050136 With Percentage of Energy Intake From Fat in Multiple Racial/Ethnic Populations. *American Journal of Epidemiology*, *178*(5), 780–790. https://doi.org/10.1093/aje/kwt028

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Pimpin, L., Ambrosini, G. L., Llewellyn, C. H., Johnson, L., van Jaarsveld, C. H., Jebb, S. A., & Wardle, J. (2013). Dietary intake of young twins: nature or nurture? *The American Journal of Clinical Nutrition*, *98*(5), 1326–1334. https://doi.org/10.3945/ajcn.113.065250

Plotnikoff, R. C., Costigan, S. A., Williams, R. L., Hutchesson, M. J., Kennedy, S. G., Robards, S. L., … Germov, J. (2015). Effectiveness of interventions targeting physical activity, nutrition and healthy weight for university and college students: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 45. https://doi.org/10.1186/s12966-015-0203-7

Pomerleau, M., Imbeault, P., Parker, T., & Doucet, E. (2004). Effects of exercise intensity on food intake and appetite in women. *The American Journal of Clinical Nutrition*, *80*(5), 1230–1236.

Prättälä, R., Paalanen, L., Grinberga, D., Helasoja, V., Kasmel, A., & Petkeviciene, J. (2007). Gender differences in the consumption of meat, fruit and vegetables are similar in Finland and the Baltic countries. *European Journal of Public Health*, *17*(5), 520–525. https://doi.org/10.1093/eurpub/ckl265

Provencher, V., Pérusse, L., Bouchard, L., Drapeau, V., Bouchard, C., Rice, T., … Lemieux, S. (2005). Familial resemblance in eating behaviors in men

and women from the Quebec Family Study. *Obesity Research*, *13*(9), 1624–1629. https://doi.org/10.1038/oby.2005.199

Qi, L., Kraft, P., Hunter, D. J., & Hu, F. B. (2008). The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Human Molecular Genetics*, *17*(22), 3502–3508. https://doi.org/10.1093/hmg/ddn242

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rankinen, T., & Bouchard, C. (2006). Genetics of Food Intake and Eating Behavior Phenotypes in Humans. *Annual Review of Nutrition*, *26*(1), 413–434. https://doi.org/10.1146/annurev.nutr.26.061505.111218

Rasmussen, M. A., & Bro, R. (2012). A tutorial on the Lasso approach to sparse modeling. *Chemometrics and Intelligent Laboratory Systems*, *119*, 21–31. https://doi.org/10.1016/j.chemolab.2012.10.003

Räty, R., & Carlsson-Kanyama, A. (2010). Energy consumption by gender in some European countries. *Energy Policy*, *38*(1), 646–649. https://doi.org/10.1016/j.enpol.2009.08.010

Reed, D. R., Bachmanov, A. A., Beauchamp, G. K., Tordoff, M. G., & Price, R. A. (1997). Heritable Variation in Food Preferences and Their Contribution to Obesity. *Behavior Genetics*, *27*(4), 373–387.

Reedy, J., Wirfält, E., Flood, A., Mitrou, P. N., Krebs-Smith, S. M., Kipnis, V., … Subar, A. F. (2010). Comparing 3 Dietary Pattern Methods—Cluster Analysis, Factor Analysis, and Index Analysis—With Colorectal Cancer Risk The NIH–AARP Diet and Health Study. *American Journal of Epidemiology*, *171*(4), 479–487. https://doi.org/10.1093/aje/kwp393

Reseland, J. E., Anderssen, S. A., Solvoll, K., Hjermann, I., Urdal, P., Holme, I., & Drevon, C. A. (2001). Effect of long-term changes in diet and exercise on plasma leptin concentrations. *The American Journal of Clinical Nutrition*, *73*(2), 240–245.

Robiou-du-Pont, S., Yengo, L., Vaillant, E., Lobbens, S., Durand, E., Horber, F., … Meyre, D. (2013). Common variants near BDNF and SH2B1 show nominal evidence of association with snacking behavior in European populations. *Journal of Molecular Medicine*, *91*(9), 1109–1115.

Ročková, V., & George, E. I. (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association*, *111*(516), 1608–1622.

Roininen, K., & Tuorila, H. (1999). Health and taste attitudes in the prediction of use frequency and choice between less healthy and more healthy snacks. *Food Quality and Preference*, *10*(4), 357–365. https://doi.org/10.1016/S0950-3293(98)00057-3

Roweis, S. T. (1998). EM algorithms for PCA and SPCA. In *Advances in neural information processing systems* (pp. 626–632).

Sachithananthan, V., & Gad, N. (2016). A Study on the Frequency of Food Consumption and Its Relationship to BMI in School Children and Adolescents in Abha City, KSA. *Current Research in Nutrition and Food Science Journal*, *4*(3), 203–208.

Saeed, S., Bonnefond, A., Tamanini, F., Mirza, M. U., Manzoor, J., Janjua, Q. M., … Froguel, P. (2018). Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nature Genetics*, *50*(2), 175–179. https://doi.org/10.1038/s41588-017-0023-6

Sailors, M. H., Jackson, A. S., McFarlin, B. K., Turpin, I., Ellis, K. J., Foreyt, J. P., … Bray, M. S. (2010). Exposing College Students to Exercise: The Training Interventions and Genetics of Exercise Response (TIGER) Study. *Journal of American College Health : J of ACH*, *59*(1), 13–20. https://doi.org/10.1080/07448481.2010.483712

Scarpace, P. J., Matheny, M., & Zhang, Y. (2010). Wheel running eliminates high-fat preference and enhances leptin signaling in the ventral tegmental area. *Physiology & Behavior*, *100*(2), 173–179. https://doi.org/10.1016/j.physbeh.2010.02.017

Schubert, M. M., Desbrow, B., Sabapathy, S., & Leveritt, M. (2013). Acute exercise and subsequent energy intake. A meta-analysis. *Appetite*, *63*, 92–104. https://doi.org/10.1016/j.appet.2012.12.010

Schulze, M. B., Hoffmann, K., Kroke, A., & Boeing, H. (2001). Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study. *British Journal of Nutrition*, *85*(03), 363. https://doi.org/10.1079/BJN2000254

Schweitzer, A. L., Ross, J. T., Klein, C. J., Lei, K. Y., & Mackey, E. R. (2016). An Electronic Wellness Program to Improve Diet and Exercise in College Students: A Pilot Study. *JMIR Research Protocols*, *5*(1). https://doi.org/10.2196/resprot.4855

Scientific Report of the 2015 Dietary Guidelines Advisory Committee. (2015). Washington (DC): Department of Health and Human Services and USDA. Retrieved from https://health.gov/dietaryguidelines/2015-scientific-report/PDFs/Scientific-Report-of-the-2015-Dietary-Guidelines-Advisory-Committee.pdf

Scott, J. G., & Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, *136*(7), 2144–2162. https://doi.org/10.1016/j.jspi.2005.08.031

Seo, D. M., Goldschmidt-Clermont, P. J., & West, M. (2007). Of mice and men: Sparse statistical modeling in cardiovascular genomics. *The Annals of Applied Statistics*, *1*(1), 152–178. https://doi.org/10.1214/07-AOAS110

Shaw, B. S., Shaw, I., & Mamen, A. (2010). Contrasting effects in anthropometric measures of total fatness and abdominal fat mass following endurance and concurrent endurance and resistance training. *The Journal of Sports Medicine and Physical Fitness*, *50*(2), 207–213.

Shaw, K., Gennat, H., O'Rourke, P., & Del Mar, C. (2006). Exercise for overweight or obesity. *The Cochrane Database of Systematic Reviews*, (4), CD003817. https://doi.org/10.1002/14651858.CD003817.pub3

Slattery, M. L., Boucher, K. M., Caan, B. J., Potter, J. D., & Ma, K.-N. (1998). Eating patterns and risk of colon cancer. *American Journal of Epidemiology*, *148*(1), 4–16.

Speakman, J. R., Rance, K. A., & Johnstone, A. M. (2008). Polymorphisms of the FTO gene are associated with variation in energy intake, but not energy expenditure. *Obesity (Silver Spring, Md.)*, *16*(8), 1961–1965. https://doi.org/10.1038/oby.2008.318

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Steinle, N. I., Hsueh, W.-C., Snitker, S., Pollin, T. I., Sakul, H., Jean, P. L. S., … Shuldiner, A. R. (2002). Eating behavior in the Old Order Amish: heritability analysis and a genome-wide linkage analysis. *The American Journal of Clinical Nutrition*, *75*(6), 1098–1106.

Steinle, N. I., Pollin, T. I., O'Connell, J. R., Mitchell, B. D., & Shuldiner, A. R. (2005). Variants in the ghrelin gene are associated with metabolic syndrome in the Old Order Amish. *The Journal of Clinical Endocrinology and Metabolism*, *90*(12), 6672–6677. https://doi.org/10.1210/jc.2005-0549

Stensel, D. (2010). Exercise, Appetite and Appetite-Regulating Hormones: Implications for Food Intake and Weight Control. *Annals of Nutrition and Metabolism*, *57*(s2), 36–42. https://doi.org/10.1159/000322702

Strong, K. A., Parks, S. L., Anderson, E., Winett, R., & Davy, B. M. (2008). Weight Gain Prevention: Identifying Theory-Based Targets for Health Behavior Change in Young Adults. *Journal of the American Dietetic Association*, *108*(10), 1708–1715. https://doi.org/10.1016/j.jada.2008.07.007

Strosberg, A. D., & Issad, T. (1999). The involvement of leptin in humans revealed by mutations in leptin and leptin receptor genes. *Trends in Pharmacological Sciences*, *20*(6), 227–230.

Stubbs, R. J., Sepp, A., Hughes, D. A., Johnstone, A. M., King, N., Horgan, G., & Blundell, J. E. (2002). The effect of graded levels of exercise on energy

intake and balance in free-living women. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity*, *26*(6), 866–869. https://doi.org/10.1038/sj.ijo.0801874

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research*, *29*(1), 71–83. https://doi.org/10.1016/0022-3999(85)90010-8

Stutzmann, F., Cauchi, S., Durand, E., Calvacanti-Proença, C., Pigeyre, M., Hartikainen, A.-L., … Froguel, P. (2009). Common genetic variation near MC4R is associated with eating behaviour patterns in European populations. *International Journal of Obesity (2005)*, *33*(3), 373–378. https://doi.org/10.1038/ijo.2008.279

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Swift, D. L., Johannsen, N. M., Lavie, C. J., Earnest, C. P., & Church, T. S. (2014). The Role of Exercise and Physical Activity in Weight Loss and

Maintenance. *Progress in Cardiovascular Diseases*, *56*(4), 441–447. https://doi.org/10.1016/j.pcad.2013.09.012

Tanofsky-Kraff, M., Han, J. C., Anandalingam, K., Shomaker, L. B., Columbo, K. M., Wolkoff, L. E., … Yanovski, J. A. (2009). The FTO gene rs9939609 obesity-risk allele and loss of control over eating12345. *The American Journal of Clinical Nutrition*, *90*(6), 1483–1488. https://doi.org/10.3945/ajcn.2009.28439

Taylor, A. H., & Oliver, A. J. (2009). Acute effects of brisk walking on urges to eat chocolate, affect, and responses to a stressor and chocolate cue. An experimental study. *Appetite*, *52*(1), 155–160. https://doi.org/10.1016/j.appet.2008.09.004

Teucher, B., Skinner, J., Skidmore, P. M. L., Cassidy, A., Fairweather-Tait, S. J., Hooper, L., … MacGregor, A. J. (2007). Dietary Patterns and Heritability of Food Choice in a UK Female Twin Cohort. *Twin Research and Human Genetics*, *10*(05), 734–748. https://doi.org/10.1375/twin.10.5.734

Thayer, R. E., Peters, D. P., Takahashi, P. J., & Birkhead-Flight, A. M. (1993). Mood and behavior (smoking and sugar snacking) following moderate exercise: A partial test of self-regulation theory. *Personality and Individual Differences*, *14*(1), 97–104. https://doi.org/10.1016/0191-8869(93)90178-6

Theobald, C., Chatterjee, A., & Horgan, G. (2012). A hierarchical Bayesian mixture model for repeated dietary records. *Food and Chemical Toxicology*, *50*(2), 320–327. https://doi.org/10.1016/j.fct.2011.10.050

Thorpe, M. G., Milte, C. M., Crawford, D., & McNaughton, S. A. (2016). A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older Australians. *The International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 30. https://doi.org/10.1186/s12966-016-0353-2

Thurstone, L. L. (1947). *Multiple-factor Analysis: A Development and Expansion of The Vectors of Mind*. University of Chicago Press.

Timpson, N. J., Emmett, P. M., Frayling, T. M., Rogers, I., Hattersley, A. T., McCarthy, M. I., & Davey Smith, G. (2008). The fat mass–and obesity-associated locus and dietary intake in children. *The American Journal of Clinical Nutrition*, *88*(4), 971–978. https://doi.org/10.1093/ajcn/88.4.971

Tipping, M. E. (2001). Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, *1*(Jun), 211–244.

Togo, P., Osler, M., Sørensen, T. I., & Heitmann, B. L. (2001). Food intake patterns and body mass index in observational studies. *International Journal of Obesity*, *25*(12), 1741.

Tucker, K. L. (2010). Dietary patterns, approaches, and multicultural perspective; Can we identify culture-specific healthful dietary patterns among diverse

populations undergoing nutrition transition? *Applied Physiology, Nutrition, and Metabolism*, *35*(2), 211–218. https://doi.org/10.1139/H10-010

Tucker, M., & Reicks, M. (2002). Exercise as a Gateway Behavior for Healthful Eating among Older Adults: An Exploratory Study. *Journal of Nutrition Education and Behavior*, *34*, S14–S19. https://doi.org/10.1016/S1499-4046(06)60306-0

U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2015). 2015–2020 Dietary Guidelines for Americans. 8th Edition. Retrieved from https://health.gov/dietaryguidelines/2015/

Vaisse, C., Clement, K., Guy-Grand, B., & Froguel, P. (1998). A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nature Genetics*, *20*(2), 113–114. https://doi.org/10.1038/2407

Valladares, M., Domínguez-Vásquez, P., Obregón, A. M., Weisstaub, G., Burrows, R., Maiz, A., & Santos, J. L. (2010). Melanocortin-4 receptor gene variants in Chilean families: association with childhood obesity and eating behavior. *Nutritional Neuroscience*, *13*(2), 71–78. https://doi.org/10.1179/147683010X12611460763643

van den Berg, L., Henneman, P., Willems van Dijk, K., Delemarre-van de Waal, H. A., Oostra, B. A., van Duijn, C. M., & Janssens, A. C. J. W. (2013).

Heritability of dietary food intake patterns. *Acta Diabetologica*, *50*(5), 721–726. https://doi.org/10.1007/s00592-012-0387-0

Varraso, R., Garcia-Aymerich, J., Monier, F., Moual, N. L., Batlle, J. D., Miranda, G., … Maccario, J. (2012). Assessment of dietary patterns in nutritional epidemiology: principal component analysis compared with confirmatory factor analysis. *The American Journal of Clinical Nutrition*, *96*(5), 1079–1092. https://doi.org/10.3945/ajcn.112.038109

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*(5), 1061–1072. https://doi.org/10.1017/S0033291714002165

Vinkhuyzen, A. A., Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2013). Estimation and Partitioning of Heritability in Human Populations using Whole Genome Analysis Methods. *Annual Review of Genetics*, *47*, 75–95. https://doi.org/10.1146/annurev-genet-111212-133258

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. https://doi.org/10.1038/nrg2322

Voight, B. F., Kang, H. M., Ding, J., Palmer, C. D., Sidore, C., Chines, P. S., … Boehnke, M. (2012). The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *PLoS Genetics*, *8*(8).

Wade, J., Milner, J., & Krondl, M. (1981). Evidence for a physiological regulation of food selection and nutrient intake in twins. *The American Journal of Clinical Nutrition*, *34*(2), 143–147. https://doi.org/10.1093/ajcn/34.2.143

Walley, A. J., Asher, J. E., & Froguel, P. (2009). The genetic contribution to non-syndromic human obesity. *Nature Reviews Genetics*, *10*(7), 431–442. https://doi.org/10.1038/nrg2594

Wang, H.-J., Yang, H.-T., & Chen, W. (2017). Swimming exercise reduces preference for a high-fat diet by increasing insulin sensitivity in C57BL/6 mice: *NeuroReport*, *28*(1), 56–61. https://doi.org/10.1097/WNR.0000000000000713

Wang, Q., Carvalho, C. M., Lucas, J., & West, M. (2007). BFRM: Software for Bayesian Factor Regression Models. *Bulleting of the International Society for Bayesian Analysis*, (14), 4–5.

Wang, Y., & Chen, X. (2011). How Much of Racial/Ethnic Disparities in Dietary Intakes, Exercise, and Weight Status Can Be Explained by Nutrition- and Health-Related Psychosocial Factors and Socioeconomic Status among

US Adults? *Journal of the American Dietetic Association*, *111*(12), 1904–1911. https://doi.org/10.1016/j.jada.2011.09.036

Wardle, J., Carnell, S., Haworth, C. M. A., Farooqi, I. S., O'Rahilly, S., & Plomin, R. (2008). Obesity Associated Genetic Variation in FTO Is Associated with Diminished Satiety. *The Journal of Clinical Endocrinology & Metabolism*, *93*(9), 3640–3643. https://doi.org/10.1210/jc.2008-0472

Wardle, J., Haase, A. M., Steptoe, A., Nillapun, M., Jonwutiwes, K., & Bellisie, F. (2004). Gender differences in food choice: The contribution of health beliefs and dieting. *Annals of Behavioral Medicine*, *27*(2), 107–116. https://doi.org/10.1207/s15324796abm2702_5

Washburn, R. A., Honas, J. J., Ptomey, L. T., Mayo, M. S., Lee, J., Sullivan, D. K., … Donnelly, J. E. (2015). Energy and macronutrient intake in the Midwest Exercise Trial-2 (MET-2). *Medicine and Science in Sports and Exercise*, *47*(9), 1941–1949. https://doi.org/10.1249/MSS.0000000000000611

Wengreen, H. J., & Moncur, C. (2009). Change in diet, physical activity, and body weight among young-adults during the transition from high school to college. *Nutrition Journal*, *8*, 32. https://doi.org/10.1186/1475-2891-8-32

West, M. (2003). Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. In *Bayesian Statistics* (pp. 723–732). Oxford University Press.

Whybrow, S., Hughes, D. A., Ritz, P., Johnstone, A. M., Horgan, G. W., King, N., … Stubbs, R. J. (2008). The effect of an incremental increase in exercise on appetite, eating behaviour and energy balance in lean men and women feeding ad libitum. *The British Journal of Nutrition*, *100*(5), 1109–1115. https://doi.org/10.1017/S0007114508968240

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.

Wilcox, S., King, A. C., Castro, C., & Bortz, W. (2000). Do changes in physical activity lead to dietary changes in middle and old age? *American Journal of Preventive Medicine*, *18*(4), 276–283.

Willett, W. C., Howe, G. R., & Kushi, L. H. (1997). Adjustment for total energy intake in epidemiologic studies. *The American Journal of Clinical Nutrition*, *65*(4), 1220S-1228S.

Williams, C. D., Satia, J. A., Adair, L. S., Stevens, J., Galanko, J., Keku, T. O., & Sandler, R. S. (2009). Dietary patterns, food groups, and rectal cancer risk in Whites and African Americans. *Cancer Epidemiology, Biomarkers & Prevention*, *18*(5), 1552. https://doi.org/10.1158/1055-9965.EPI-08-1146

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Visscher, P. M. (2010). Common SNPs explain a large proportion of heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., … Visscher, P. M. (2011). Genome-partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, *43*(6), 519–525. https://doi.org/10.1038/ng.823

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*(9), 1304–1310. https://doi.org/10.1038/ng.3941

Yang, T., Xu, W.-J., York, H., & Liang, N.-C. (2017). Diet choice patterns in rodents depend on novelty of the diet, exercise, species, and sex. *Physiology & Behavior*, *176*, 149–158. https://doi.org/10.1016/j.physbeh.2017.02.045

Yeo, G. S. H., Farooqi *, I. S., Aminian, S., Halsall, D. J., Stanhope, R. G., & O'Rahilly, S. (1998). A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genetics*, *20*(2), 111–112. https://doi.org/10.1038/2404

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., & Friedman, J. M. (1994). Positional cloning of the mouse obese gene and its human

homologue. *Nature*, *372*(6505), 425–432. https://doi.org/10.1038/372425a0

Zorrilla, E. P., Inoue, K., Valdez, G. R., Tabarin, A., & Koob, G. F. (2005). Leptin and post-prandial satiety: acute central leptin more potently reduces meal frequency than meal size in the rat. *Psychopharmacology*, *177*(3), 324–335. https://doi.org/10.1007/s00213-004-1952-1

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://doi.org/10.1198/106186006X113430